December 2016

Riboregulation of Bacterial Transposons

Michael J. Ellis

The University of Western Ontario

Supervisor

David Haniford

The University of Western Ontario

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Michael J. Ellis 2016

Follow this and additional works at: http://ir.lib.uwo.ca/etd

Part of the Molecular Biology Commons

Recommended Citation

http://ir.lib.uwo.ca/etd/4280

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca.
Abstract

Bacterial transposons typically exist in a mutually beneficial relationship with the host cell. Limited transposition can benefit the host while also ensuring the survival of the element. An important component of this relationship is that transposition must be tightly regulated. In this thesis I explore ways that the host and transposon each control transposition in E. coli and provide evidence that a transposon can also control host gene expression in S. enterica Typhimurium. Post-transcriptional regulation with small non-coding RNAs (sRNA) has emerged as a key way that bacteria respond to stress and regulate many cellular processes. The RNA-binding protein Hfq is the nexus of sRNA regulatory networks and acts by promoting base-pairing interactions between sRNAs and their target mRNAs. Previous work found that Hfq is a potent negative regulator of IS10 transposition in E. coli and suggested that Hfq inhibited transposase translation using an IS10-encoded sRNA (RNA-OUT) as well as an undefined mechanism that was independent of RNA-OUT. I show that Hfq promotes base-pairing between RNA-OUT and IS10 transposase mRNA (RNA-IN) by melting the secondary structure of both RNAs to expose residues involved in intermolecular base-pairing. I also investigated how Hfq can repress translation of RNA-IN in the absence of RNA-OUT and demonstrate that Hfq-binding to an mRNA can directly repress translation in the absence of any sRNA. The data suggested Hfq may regulate other transposons and I show that the unrelated IS200 element is also subject to Hfq regulation. In contrast to the IS10 system, Hfq represses IS200 transposase (tnpA) translation completely independent of the IS200-encoded sRNA (art200). Translation initiation on tnpA is inhibited >350-fold by the cooperation of Hfq, art200, and an RNA structural element in the tnpA 5’UTR illustrating how host- and transposon-encoded factors can coordinate to repress transposition. Lastly, I demonstrate that tnpA is processed to produce an sRNA that alters transcript abundance >2-fold for 73 S. enterica Typhimurium genes, which provides a new twist on our understanding of host-transposon interactions. Taken together, this work suggests that RNA transactions play an important role in governing host-transposon relationships in bacteria.
Keywords

Co-Authorship Statement

For the peer-reviewed publication presented in Chapter 2 J. Ross and M. Ellis were co-first authors. J. Ross performed the work presented in Figures 2.2, 2.6-2.8, and S2.4-S2.7 with the assistance of S. Hossain. J. Ross and D. Haniford drafted the manuscript with the exception of sections 2.2.2, 2.3.3, and 2.4.4 which were written by M. Ellis and D. Haniford. Revisions to the manuscript were made by J. Ross, M. Ellis, and D. Haniford.

For the peer-reviewed publications presented in Chapters 3, 4, and 5 (In Revision), M. Ellis and D. Haniford conceived and designed the studies and wrote the manuscripts. M. Ellis performed the experimental work with the exceptions noted below. M. Ellis analyzed the data and prepared figures.

Chapter 3 – D. Haniford performed one of the mating experiments presented in Figure 3.2 and the sRNA screen presented in Figure 3.8. R. Trussler performed primer extension experiments presented in Figure 3.8 and 3.9, as well as the Miller Assay in Figure S3.6. D. Haniford constructed strains.

Chapter 4 – R. Trussler was a co-first author and performed the work presented in Figures 4.2, 4.3, 4.6, S4.1, and S4.5-4.7. R. Trussler performed the RNase footprinting in Figure 4.4, the Miller Assay in Figure 4.7, and the Southern Blot in Figure 4.9. D. Haniford performed the mating out assays in Figure 4.9. R. Trussler and D. Haniford constructed strains and plasmids.

Chapter 5 – R. Trussler performed the work presented in Figures 5.3, S5.2, and S5.5. R. Trussler and D. Haniford constructed strains and plasmids.
Acknowledgments

The most important acknowledgment for this thesis is owed to my supervisor, Dr. David Haniford, who has given me a tremendous amount of guidance and support throughout my Doctoral work. As a supervisor, you helped me develop into a critical and independent scientist and I sincerely appreciate your patience and willingness to let me try new things even when it must have seemed like I was on wild goose chase. You have instilled in me the importance of treating data fairly and ‘getting it right’, and for this I thank you.

I want to thank the past and present members of my advisory committee—Dr. Dave Edgell, Dr. Chris Brandl, and Dr. Megan Davey—for the advice and technical guidance you have given me over the past 5 years. In particular, I’m grateful to Dave and Chris for always taking time to answer ‘one quick question’. My thanks go out to other members of the Biochemistry Department: Barb Green, Dr. Murray Junop, Dr. David Litchfield, Dr. Greg Gloor, Dr. Stan Dunn, Yumin Bi, and others.

To the past and present members of the Haniford lab—especially Ryan, Joe, Brian, Morgan, Shahan, and Crystal—thank you for all of the great conversations, moral support, tolerating my persistent optimism, shirt tans, and in general making the last 5 years so enjoyable. To Joe, thanks for the many arguments (often about science) and ridiculous conversations, and for taking me under your wing when I started. To Ryan, I can’t thank you enough for all of the work you have done, for acting not just as a sounding board but an invaluable critic, and for been a great lab mate for the last 5 years. I also want to thank members of the Brandl lab (Julie, Sam, Matt, Kyle, and Matt) for being great next-door neighbors.

Finally, I want to thank my amazing wife Katie and beautiful daughter Sophie. Katie, your unending support and motivation is the reason I am where I am today. The last 5 years has probably seemed like an eternity but coming home to the both of you has always made the difficult days better and the good days great. I can’t wait to face the future with you by my side.
Table of Contents

Abstract ................................................................................................................................. i
Co-Authorship Statement ..................................................................................................... iii
Acknowledgments ................................................................................................................ iv
Table of Contents ................................................................................................................ v
List of Tables ........................................................................................................................ xi
List of Figures ....................................................................................................................... xii
List of Supplemental Figures .............................................................................................. xv
List of Supplemental Tables ................................................................................................. xvii
List of Appendices ................................................................................................................ xviii
List of Abbreviations and Symbols ..................................................................................... xix

Chapter 1 ............................................................................................................................ 1

1 General Introduction ......................................................................................................... 1

1.1 Introduction to bacterial transposons ........................................................................... 2

1.2 Impact of transposons on genome evolution and adaptation ..................................... 6

1.3 Regulation of bacterial transposons ............................................................................ 8

1.3.1 Intrinsic control of transposition .......................................................................... 8

1.3.2 Host factors affecting transposition ..................................................................... 10

1.4 Transposons and stress response ................................................................................ 11

1.5 Regulatory RNA in bacteria ....................................................................................... 12

1.5.1 Base-pairing small RNAs (sRNA) ...................................................................... 12

1.5.2 Antisense RNAs (asRNA) ................................................................................ 17

1.5.3 RNAs that regulate protein activity and riboswitches ....................................... 19

1.6 Hfq .............................................................................................................................. 19

1.6.1 Structure and RNA-binding properties ............................................................ 20
1.6.2 Role of Hfq in post-transcriptional regulation .................................................. 23

1.7 Mobile genetic elements and pathogenesis of *Salmonella* Typhimurium ........ 26

1.8 Scope of this thesis ...................................................................................................... 27

1.9 References .................................................................................................................. 29

Chapter 2 .......................................................................................................................... 43

2 Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the
Tn10/IS10 system ............................................................................................................. 43

2.1 Introduction .................................................................................................................. 43

2.2 Results .......................................................................................................................... 46

2.2.1 Hfq binds RNA-IN approximately 80-fold more tightly than RNA-OUT ......... 46

2.2.2 RNase and hydroxyl radical footprinting of RNA-IN, RNA-OUT, and
Hfq:RNA complexes ........................................................................................................ 49

2.2.3 RNA-binding sites in Hfq that interact with RNA-IN and RNA-OUT ..... 59

2.2.4 Hfq\textsubscript{K56A} exhibits a reduced rate of RNA-IN:RNA-OUT pairing *in vitro* . 66

2.2.5 IS10-kan transposition is derepressed in strains expressing Y25A and
K56A forms of Hfq .......................................................................................................... 68

2.3 Discussion ................................................................................................................... 70

2.3.1 RNA-IN and RNA-OUT bind Hfq like a prototypical sRNA-mRNA pair
......................................................................................................................................... 70

2.3.2 A proximal site mutation impacts on RNA-IN:OUT pairing and IS10
transposition ...................................................................................................................... 72

2.3.3 How might Hfq promote pairing in the IS10 antisense system? ................. 73

2.4 Materials and Methods ............................................................................................ 77

2.4.1 Strains, plasmids, and primers ............................................................................. 77

2.4.2 Hfq purification and quantitation .......................................................................... 77

2.4.3 *In vitro* transcription and RNA purification ......................................................... 80

2.4.4 Hfq-RNA binding assays ....................................................................................... 80

2.4.5 RNA structure-probing and footprinting ............................................................... 82
Chapter 2

2.4.6 Determination of RNA-OUT:RNA-IN pairing rates ........................................... 83

2.4.7 Mating out assay ................................................................................................. 84

2.5 Supplemental Material .......................................................................................... 85

2.6 References ............................................................................................................. 97

Chapter 3

3 Hfq binds directly to the ribosome binding site of IS10 transposase mRNA to inhibit translation ......................................................................................................................... 101

3.1 Introduction .............................................................................................................. 101

3.2 Results ..................................................................................................................... 106

3.2.1 Antisense-independent regulation of transposase expression requires the distal surface of Hfq ................................................................. 106

3.2.2 The distal surface of Hfq binds the RBS of RNA-IN in vitro ...................... 108

3.2.3 Hfq binding to site 1 is critical for regulation of transposase expression in vivo .......................................................................................... 112

3.2.4 Hfq blocks 30S ribosome binding to RNA-IN in vitro .......................... 116

3.2.5 Hfq binds native RNA-IN in vivo and this binding is inhibited by site 1 mutations ......................................................................................... 119

3.2.6 ChiX overexpression titrates Hfq away from RNA-IN ......................... 121

3.3 Discussion .............................................................................................................. 127

3.3.1 Direct repression of IS10 translation by Hfq .............................................. 127

3.3.2 Hfq regulation of IS10 in single and multi-copy ...................................... 129

3.3.3 Hfq titration by the sRNA ChiX .................................................................. 132

3.4 Materials and Methods ......................................................................................... 133

3.4.1 Bacterial strains, phage, plasmids and oligonucleotides ......................... 133

3.4.2 Hfq footprinting and EMSA ....................................................................... 135

3.4.3 β-galactosidase assays .................................................................................. 136

3.4.4 Conjugal mating out assay .......................................................................... 136

3.4.5 RNA extraction and primer extension analysis ....................................... 136
3.4.6 Toeprinting................................................................................................. 137
3.4.7 Hfq-RNA immunoprecipitation (RIP)..................................................... 138
3.4.8 RT-PCR .................................................................................................... 139
3.4.9 Northern Blot .......................................................................................... 139
3.5 Supplemental Material ............................................................................... 139
3.6 References................................................................................................... 153
Chapter 4 ......................................................................................................... 158

4 A cis-encoded sRNA, Hfq and mRNA secondary structure act independently to suppress IS200 transposition ........................................................................ 158

4.1 Introduction ............................................................................................... 158

4.2 Results........................................................................................................ 163

4.2.1 Characterization of the IS200 antisense RNA gene in E. coli ........... 163
4.2.2 The IS200 antisense RNA inhibits IS200 transposase expression .... 166
4.2.3 Hfq negatively regulates tnpA expression but independent of art200 .. 169
4.2.4 art200 and tnpA mRNA interact in vitro ............................................. 170
4.2.5 Base-pairing between art200 and tnpA blocks 30S ribosome binding in vitro and inhibits transposase expression in vivo ....................... 173
4.2.6 tnpA translation is also repressed by mRNA secondary structure .... 178
4.2.7 Hfq blocks ribosome binding to tnpA in vitro ................................... 180
4.2.8 IS200 transposition is limited by translational control .................... 182

4.3 Discussion ................................................................................................. 186

4.3.1 Translation of the IS200 transposase is repressed by a cis-encoded sRNA, Hfq and RNA secondary structure ........................................ 186
4.3.2 Might art200 function as both a cis and trans acting sRNA? ......... 190
4.3.3 IS200 5’UTR as a platform for designing novel riboregulators of translation initiation .............................................................. 191

4.4 Materials and Methods ........................................................................... 192

4.4.1 Bacterial strains, plasmids and oligonucleotides ............................... 192
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4.2</td>
<td>RNA footprinting and toeprinting</td>
<td>194</td>
</tr>
<tr>
<td>4.4.3</td>
<td>β-galactosidase assays</td>
<td>195</td>
</tr>
<tr>
<td>4.4.4</td>
<td>RNA extraction, primer extension</td>
<td>195</td>
</tr>
<tr>
<td>4.4.5</td>
<td>Conjugal mating out assay</td>
<td>195</td>
</tr>
<tr>
<td>4.5</td>
<td>Supplemental Material</td>
<td>196</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Materials and Methods</td>
<td>196</td>
</tr>
<tr>
<td>4.6</td>
<td>Supplemental Figures</td>
<td>199</td>
</tr>
<tr>
<td>4.7</td>
<td>References</td>
<td>215</td>
</tr>
<tr>
<td>Chapter 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>A transposon derived small RNA regulates gene expression in Salmonella Typhimurium</td>
<td>222</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>222</td>
</tr>
<tr>
<td>5.2</td>
<td>Results</td>
<td>226</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Profiling changes in S. Typhimurium gene expression in response to altered levels of IS200-encoded transcripts</td>
<td>226</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Characterization of tnpA derived RNAs</td>
<td>228</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Over-expression of tnpA represses expression of SPI-1 in a growth phase dependent manner</td>
<td>234</td>
</tr>
<tr>
<td>5.2.4</td>
<td>Direct interaction between tnpA and invF</td>
<td>238</td>
</tr>
<tr>
<td>5.3</td>
<td>Discussion</td>
<td>240</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Ribonucleolytic processing of tnpA mRNA generates an sRNA regulator of invF expression</td>
<td>240</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Regulatory cross-talk between horizontally acquired genes and the S. Typhimurium core genome</td>
<td>243</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Bacterial transposons as a source of regulatory RNA</td>
<td>246</td>
</tr>
<tr>
<td>5.4</td>
<td>Materials and Methods</td>
<td>246</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Growth conditions, strains, and plasmids</td>
<td>246</td>
</tr>
<tr>
<td>5.4.2</td>
<td>RNA isolation, Northern blot, and primer extension</td>
<td>247</td>
</tr>
</tbody>
</table>
5.4.3 RNA-Seq and data analysis ................................................................. 247
5.4.4 Reverse transcription quantitative polymerase chain reaction (RT-qPCR)
.............................................................................................................. 248
5.4.5 Western Blot .................................................................................... 249
5.4.6 Electrophoretic Mobility Shift Assay (EMSA) and Lead Footprinting.. 249
5.5 Supplemental Material ........................................................................ 250
  5.5.1 Supplemental Materials and Methods ........................................... 250
  5.5.2 Supplemental Tables and Figures .................................................... 252
5.6 References ......................................................................................... 263

Chapter 6 .................................................................................................. 268
6 General Discussion .................................................................................. 268
  6.1 Host-transposon interactions at the RNA level .................................... 269
    6.1.1 Hfq represses IS10 transposase expression by two distinct mechanisms 269
    6.1.2 Cooperation between host- and transposon-encoded factors controls
        transposase translation ........................................................................ 274
    6.1.3 Biogenesis of a trans-acting sRNA from a transposase mRNA ....... 275
  6.2 An integrated model for repression of IS10/Tn10 transposition by Hfq .... 277
  6.3 Is post-transcriptional regulation a general mechanism employed by host bacteria
        to silence transposons? ................................................................. 281
  6.4 Concluding remarks .......................................................................... 282
  6.5 References ....................................................................................... 282

Appendices ............................................................................................... 287
Curriculum Vitae ...................................................................................... 292
List of Tables

Table 2.1. *In vitro* binding of RNA-OUT or RNA-IN to WT, distal-, or proximal-impaired Hfq mutants........................................................................................................................................ 48

Table 2.2. *In vitro* competition by DsrA or A18 for binding of HfqWT to RNA-IN or RNA-OUT ........................................................................................................................................... 64

Table 2.3. List of *E. coli* strains and plasmids used in Chapter 2....................................................... 78

Table 2.4. List of oligonucleotides used in Chapter 2. ....................................................................... 79

Table 4.1. Strains and plasmids used in Chapter 4. .......................................................................... 193
List of Figures

Figure 1.1. Overview of bacterial transposons ........................................................... 4
Figure 1.2. Common mechanisms for sRNA-mediated gene regulation ..................... 14
Figure 1.3. Common mechanisms employed by antisense RNAs .............................. 18
Figure 1.4. Structure of Hfq .................................................................................. 22
Figure 1.5. Main functions of Hfq in post-transcriptional regulation ...................... 24
Figure 2.1. Small regulatory RNAs (sRNAs) and the Tn10/IS10 antisense system ........ 44
Figure 2.2. Hfq binds with high and moderate affinities to RNA-IN and RNA-OUT in vitro. ................................................................................................................. 47
Figure 2.3. Structure-probe analysis of RNA-OUT and Hfq:RNA-OUT complex........ 50
Figure 2.4. RNase footprinting of RNA-IN ............................................................. 54
Figure 2.5. Hydroxyl radical footprinting of RNA-IN ............................................. 58
Figure 2.6. RNA-OUT interacts specifically with the proximal RNA-binding surface of Hfq. ......................................................................................................................... 60
Figure 2.7. RNA-IN interacts with the distal and proximal RNA-binding surfaces of Hfq... 65
Figure 2.8. RNA-IN:RNA-OUT pairing reactions .................................................. 67
Figure 2.9. IS10-kan transposition is derepressed in E. coli encoding Y25A and K56A forms of Hfq ................................................................................................................. 69
Figure 2.10. Model for RNA-IN:OUT antisense pairing in the presence vs. absence of Hfq. 74
Figure 3.1. Overview of the Tn10/IS10 system .......................................................... 104
Figure 3.2. Impact of mutant forms of Hfq on IS10 transposase expression and transposition. .............................................................................................................................................................................................. 107

Figure 3.3. RNA-IN footprinting with WT and mutant forms of Hfq. .................................................. 110

Figure 3.4. Effect of RNA-IN mutations on Hfq binding. ................................................................ 113

Figure 3.5. Impact of mutant forms of RNA-IN on transposase expression in hfq''/hfq'' strains. .................................................................................................................................. 115

Figure 3.6. Impact of Hfq on initiation of RNA-IN translation in vivo. ........................................... 117

Figure 3.7. Hfq-RNA-IN immunoprecipitation (RIP) assay. .............................................................. 120

Figure 3.8. ChiX positively regulates IS10 transposase translation. ................................................ 123

Figure 3.9. ChiX competes with RNA-IN for Hfq-binding in vivo. ...................................................... 126

Figure 3.10. Model for dual role of Hfq in repressing IS10 transposase translation................. 131

Figure 4.1. Schematic of IS200........................................................................................................... 161

Figure 4.2. Primer extension analysis of art200 in E. coli. ................................................................. 165

Figure 4.3. Impact of art200 and Hfq on tnpA expression................................................................. 167

Figure 4.4. Pb^{2+} and RNase footprinting of art200, tnpA_{1-73} and an art200-tnpA_{1-173} complex. ........................................................................................................................................... 172

Figure 4.5. Impact of antisense pairing on ribosome binding to tnpA_{1-73} in vitro. .......... 175

Figure 4.6. Impact of terminal loop mutations on translational repression in vivo. .......... 177

Figure 4.7. Impact of tnpA lower stem (LS) mutations on transposase expression........ 179

Figure 4.8. Hfq inhibits 30S ribosomal subunit binding to tnpA and binds upstream of the SD. ........................................................................................................................................... 181

Figure 4.9. IS200 transposition assays.............................................................................................. 184
Figure 4.10. Model for translational repression of IS200 \textit{tnpA}. .................................................. 189

Figure 5.1. IS200 and experimental approach. ................................................................. 224

Figure 5.2. Summary of RNA-Seq data.............................................................................. 227

Figure 5.3. Processing of the \textit{tnpA} transcript.................................................................... 229

Figure 5.4. The 5'end of \textit{tnpA} represses \textit{invF} mRNA and protein expression.............. 233

Figure 5.5. Overexpression of \textit{tnpA} RNA downregulates \textit{invF} and \textit{sicAsipBC} in a growth phase dependent manner. ........................................................................ 235

Figure 5.6. Reducing IS200 copy-number increases \textit{invF} expression.............................. 237

Figure 5.7. Evidence for a base-pairing interactions between \textit{invF} and \textit{tnpA}. ............... 239

Figure 5.8. Model for role of \textit{tnpA} in regulating SPI-1 gene expression....................... 245

Figure 6.1. Summary of new insights into post-transcriptional control of transposons IS10 and IS200 gained in this work. ........................................................................ 271

Figure 6.2. Influence of Dam methylation and transposon copy-number on Hfq-mediated repression of IS10 transposition........................................................................ 279
List of Supplemental Figures

Figure S 2.1. RNA footprinting EMSAs................................................................. 85
Figure S 2.2. RNA-IN-160 RNase footprinting...................................................... 87
Figure S 2.3. Structure of RNA-OUT and RNA-IN in the presence of Hfq.............. 89
Figure S 2.4. IC50 determinations for A18 and/or DsrA competition experiments.......... 91
Figure S 2.5. Measurement of RNA-OUT:RNA-IN pairing as a function of Hfq concentration...................................................................................................................... 92
Figure S 2.6. Hfq western blots. ................................................................................ 93
Figure S 2.7. RNA-OUT half-life analysis in isogenic hfq+/- backgrounds............... 95
Figure S 3.1. Hfq expression in an hfq strain of E. coli (DBH299) used for complementation experiments...................................................................................................................... 145
Figure S 3.2. Mutations to the ribosome-binding site (RBS) of RNA-IN prevent Hfq binding. ............................................................................................................................... 146
Figure S 3.3. Toeprint analyis of 30S ribosome binding to lpp or usg mRNA in the presence of Hfq.......................................................... 147
Figure S 3.4. Specificity of Hfq:RNA-IN IP and size distribution of RNA-IN molecules containing a 5’end................................................................. 148
Figure S 3.5. Hfq determinants for ChiX binding. ................................................. 151
Figure S 3.6. Impact of mRNA overexpression on transposase expression........ 152
Figure S 4.1. Impact of Hfq on tnpA-lacZ transcript levels...................................... 199
Figure S 4.2. Lead and RNase structure probing of art200. .................................... 200
Figure S 4.3. RNase structure probing of tnpA1-173................................................. 202
Figure S 4.4. Hfq blocks ribosome binding to *tnpA in vitro*............................. 204
Figure S 4.5. Schematic of conjugal “mating out” assay for IS200 transposition............. 205
Figure S 4.6. Validation of IS200 transposition events (‘hops’). .................................... 207
Figure S 4.7. Impact of mutations to a putative translational enhancer in *tnpA* on *tnpA* expression. ........................................................................................................ 209
Figure S 4.8. Predicted alternative secondary structures of the first 120nt of *tnpA*. ........ 210
Figure S 5.1. Growth phase dependent regulation of art200 and art200 knockdown by *tnpA* trunc overexpression. ........................................................................................................ 256
Figure S 5.2. Processing of *tnpA*.................................................................................. 257
Figure S 5.3. Schematic of *tnpA* overexpression constructs used in this work. .............. 259
Figure S 5.4. Growth phase dependent regulation of *invF* by *tnpA*............................ 260
Figure S 5.5. The T1 mutations to *tnpA* disrupt base-pairing with *invF in vitro*......... 262
List of Supplemental Tables

Table S 3.1. Strains, plasmids, and bacteriophage for Chapter 3 ........................................ 140

Table S 3.2. List of Oligonucleotides for Chapter 3 .............................................................. 142

Table S 4.1. Bacteriophage and supplemental plasmids for Chapter 4. ................................. 212

Table S 4.2. List of oligonucleotides for Chapter 4................................................................. 213

Table S 5.1. Strains and plasmids for Chapter 5................................................................. 253

Table S 5.2. List of oligonucleotides for Chapter 5......................................................... 254
List of Appendices

Appendix A. Copyright Permissions. ................................................................................. 287
List of Abbreviations and Symbols

ap  ampicillin
arg  arginine
asRNA  antisense RNA
bp  base-pair
cf.  compare
clr  centered log ratio
cm  chloramphenicol
CT  cycle threshold
CTD  C-terminal domain
ddH₂O  double-distilled water
DEPC  diethylpyrocarbonate
DNA  deoxyribonucleic acid
DNase  deoxyribonuclease
dNTP  deoxynucleoside triphosphate
dsRNA  double-strand RNA
DTT  dithiothreitol
EMSA  electrophoretic mobility shift assay
FRT  flippase recognition target
His  histidine
hr  hour
IE  inside end
IP  immunoprecipitation
IPTG  isopropyl β-D-1-thiogalactopyranoside
IS  insertion sequence
ITR  in vitro transcription
Kan  kanamycin
Kb  kilobase-pair
K_{D}  dissociation constant
kDa  kilodalton
leu  leucine
LS  lower stem
MCI  multi-copy inhibition
MGE  mobile genetic element
Min  minute
mRNA  messenger RNA
ncRNA  non-coding RNA
nt  nucleotide
OD  optical density
OE  outside end
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>Pro</td>
<td>proline</td>
</tr>
<tr>
<td>pTet</td>
<td>tetracycline promoter</td>
</tr>
<tr>
<td>RBS</td>
<td>ribosome binding site</td>
</tr>
<tr>
<td>RIP</td>
<td>Hfq-RNA immunoprecipitation</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>RNasin</td>
<td>ribonuclease inhibitor</td>
</tr>
<tr>
<td>rNTP</td>
<td>ribonucleoside triphosphate</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>RT</td>
<td>reverse transcription</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcription polymerase chain reaction</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>reverse transcription quantitative polymerase chain reaction</td>
</tr>
<tr>
<td>SCV</td>
<td><em>Salmonella</em> containing vacuole</td>
</tr>
<tr>
<td>SD</td>
<td>Shine-Dalgarno sequence</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>Sm</td>
<td>streptomycin</td>
</tr>
<tr>
<td>SPI</td>
<td><em>Salmonella</em> pathogenicity island</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>sRNA</td>
<td>small RNA</td>
</tr>
<tr>
<td>SSC</td>
<td>saline sodium citrate</td>
</tr>
<tr>
<td>ssRNA</td>
<td>single-strand RNA</td>
</tr>
<tr>
<td>ST-PCR</td>
<td>semi-random two-step PCR</td>
</tr>
<tr>
<td>T3SS</td>
<td>type-three secretion system</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate EDTA</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris-borate EDTA</td>
</tr>
<tr>
<td>TE</td>
<td>Translational enhancer</td>
</tr>
<tr>
<td>Tet</td>
<td>tetracycline</td>
</tr>
<tr>
<td>TEX</td>
<td>5’monophosphate dependent terminator exonuclease</td>
</tr>
<tr>
<td>TIR</td>
<td>translation initiation region</td>
</tr>
<tr>
<td>TLF</td>
<td>translational fusion</td>
</tr>
<tr>
<td>Tn</td>
<td>transposon</td>
</tr>
<tr>
<td>Tnp</td>
<td>transposase</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TSS</td>
<td>transcription start site</td>
</tr>
<tr>
<td>UT</td>
<td>untreated</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>WT</td>
<td>wild-type</td>
</tr>
</tbody>
</table>
$\sigma$  sigma factor

$\sim$  approximately

$\lambda$  bacteriophage lambda
Chapter 1

1 General Introduction

Transposable elements are ubiquitous in nature and have had a profound impact on reshaping genomes throughout evolution. The radical idea that genes could ‘jump’ throughout a genome and change chromosome structure and gene expression was first proposed almost 70 years ago by Barbara McClintock (McClintock, 1948). McClintock was studying the genetic determinants of kernel colour on the short arm of chromosome 9 in Zea mays (maize) when she observed an increase in unstable mutations (‘mutable loci’) that led to phenotypic mosaicism (i.e. variegation) on individual kernels. By studying the chromosomes of different parts of a single kernel (which arose from a single zygote), McClintock found that the changes in phenotype were due to physical changes in the chromosome and this provided the first evidence of a dynamic genome. These mutable loci were proposed to be caused by the action of two unlinked genetic elements, Ac (Activator) and Ds (Dissociator). The Ac element caused double-strand breaks in chromosome 9 that led to movement, or transposition, of Ac or Ds DNA to new locations in the genome. Ac/Ds transposition led to fully reversible changes in gene expression at the loci they moved to, as well as chromosomal abnormalities that resulted from breakage (McClintock, 1948; McClintock, 1949; McClintock, 1950). It took the wider scientific community almost 30 years to accept McClintock’s concept of a dynamic genome, where pieces of DNA could physically move throughout a genome and alter gene expression of the ‘host’ organism.

Transposons have now been identified in all domains of life (Biémont, 2010; Filee et al, 2007; Siguier et al, 2014). Despite being initially viewed as ‘junk’ or ‘selfish’ DNA, transposons are now recognized as a driving force of evolution and genome plasticity in both eukaryotes and prokaryotes (Feschotte, 2008; Feschotte & Pritham, 2007; Lapp & Hunter, 2016; Patel, 2016; Siguier et al, 2014; Volff, 2006). The focus of this thesis is to explore the relationship between bacterial transposons and the host cell. Specifically, I have investigated the ways in which the host organism and transposon together regulate transposition to minimize the burden of these potentially mutagenic genetic elements. Beginning with the model transposon, Tn10/IS10, I show in Chapters 2
and 3 that the *Escherichia coli* RNA-binding protein Hfq regulates expression of the transposase protein by two distinct mechanisms. In Chapter 4 I extend these studies to a poorly understood transposon, IS200, and characterize a combination of transposon- and host-encoded factors that work synergistically to silence transposition. Lastly, I present work in Chapter 5 that shows how a bacterial transposon can affect virulence in *Salmonella enterica* subsp. *enterica* serovar Typhimurium (*Salmonella* Typhimurium) by a previously unknown mechanism.

### 1.1 Introduction to bacterial transposons

It took 15 years from McClintock’s discovery of *Ac/Ds* elements in maize before the first transposable element was discovered in bacteria. Infection of *E. coli* with a temperate bacteriophage led to a mutator phenotype where the bacteriophage could move throughout the genome and affect expression of the genes at or near the integration site. This bacteriophage was named Mu and provided the first example of replicative transposition (Taylor, 1963). Soon after, small fragments of foreign DNA causing polar mutations in the *gal* operon and bacteriophage λ were discovered and termed Insertion Sequences (IS) (Fiandt et al, 1972; Hirsch et al, 1972a; Hirsch et al, 1972b; Shapiro, 1969). Surprisingly, these IS elements were found to be a normal component of the *E. coli* genome and often present in multiple copies (Saedler & Heiss, 1973). Similar to *Ac/Ds* and Mu, transposition of IS elements could alter expression of genes located at the insertion site (Saedler et al, 1974). The discovery of IS elements was soon followed by the observation that antibiotic resistance genes could move between plasmids and the chromosome (Barth et al, 1976; Hedges & Jacob, 1974), and these elements were first named transposons (Tn). The relationship between Tn and IS elements was finally recognized when two inverted copies of IS1 were found flanking a transposable toxin in Enterotoxigenic *E. coli* (ETEC) (So et al, 1979).

Throughout this thesis I will use the generic term transposon to refer to all transposable elements. However, there is tremendous variety in the genetic organization, size, and mechanism of bacterial transposons. The simplest transposons are IS elements,
which encode only the transposase enzyme (Tnp) required for transposition reactions. The transposase gene (\textit{tnp}) is typically flanked by inverted repeats which are recognized by Tnp as the boundaries of the transposon. Composite transposons (Tn) consist of two IS elements flanking extra genes (passenger genes) that are not involved in transposition; these passenger genes often provide a selective advantage to the host organism (e.g. determinants for antibiotic resistance or virulence). Only one copy of the IS is required to produce a functional Tnp, which can catalyze transposition of the entire Tn or only the IS (Figure 1.1). The classic distinction between IS and Tn is complicated by the discovery of IS that carry passenger genes, conjugative Tn that also encode determinants for intercellular transfer, and non-autonomous transposons that require a Tnp expressed from an unlinked transposon (Siguier et al, 2015).
Figure 1.1. Overview of bacterial transposons.

(A) Comparison of Tn and IS elements. Composite transposons (Tn) consist of two insertion sequences (IS; orange) flanking extra genes (passenger genes, blue) which are not involved in transposition reactions and typically provide a selective advantage to the host (e.g. antibiotic resistance genes). IS elements consist of a transposase gene (light orange) flanked by transposon ends (dark orange) which almost always contain inverted repeat sequences. Flanking ‘host’ DNA is shown in green. Note that the two IS elements in a Tn do not necessarily have to be identical, and only one copy needs to encode a functional transposase protein; IS elements in a Tn can also transpose independent of the rest of the Tn. (B) General transposition reaction for a non-replicative transposon (e.g. IS10). The transposase protein (blue) is expressed from an IS element; the transposase promoter is often located within or near the inverted repeats (i). Two monomers of transposase bind the transposon ends (dark orange, ii) to form a transpososome. A transposase heterodimer catalyzes the excision of transposon DNA (iii) from the donor site (light green), and then catalyzes the insertion of the transposon DNA into a new
DNA sequence (dark green, iv). The transpososome is disassembled following insertion and the transposon is now stably integrated into a new locus in the chromosome (v). Note that no genetic material is lost at the insertion site but many transposons cause a target site duplication.
1.2 Impact of transposons on genome evolution and adaptation

Transposons were initially dismissed as selfish elements that propagate at the expense of the host organism. Transposition events can cause double-strand breaks, insertions, deletions, and other chromosomal rearrangements that all contribute to genomic instability. Accordingly, they were viewed as molecular parasites. However, transposons contribute to the rapid evolution of bacterial genomes and can provide a selective advantage to the host (Lee et al., 2016; Nevers & Saedler, 1977; Siguier et al., 2014; Sundaram et al., 2014). In this way, transposons exist in a mutually beneficial relationship with the host bacterium, where limited transposition can benefit the host as well as ensuring survival of the element.

One broad way in which transposons affect the host is by facilitating large chromosomal rearrangements. Transposon expansion through repeated transposition generates homologous sequences dispersed throughout a genome. Multi-copy transposons are then sites of intrachromosomal recombination that results in large-scale inversions, deletions, and rearrangements (Alokam et al., 2002; Daveran-Mingot et al., 1998; Haack & Roth, 1995; He et al., 2015; Lee et al., 2016; Mahillon & Chandler, 1998). As bacteria are asexual, transposon-mediated genome rearrangements are major drivers of evolution. Transposon expansion is also thought to be the first step in genome streamlining that many host-adapted pathogens undergo (Plague et al., 2008; Schmitz-Esser et al., 2011; Siguier et al., 2014; Touchon & Rocha, 2007). In addition to generating chromosomal rearrangements and deletions, increased transposition promotes the formation of pseudogenes that are subsequently lost due to purifying selection (Mira et al., 2001; Moran & Plague, 2004). Transposition typically generates more severe mutations than point mutations and therefore accelerates the process of pseudogene formation that precedes genome reduction.

Transposons can also affect gene expression by shuffling regulatory elements or inserting into the genome such that genes are turned on or off. In Burkholderia cepacia, insertion of IS1490 creates a fusion promoter for the tflAB gene cluster and accordingly increases expression of catabolic genes for the toxin 2,4,5-trichlorophenoxyacetic acid,
allowing cells to use this herbicide as a sole carbon source (Hubner & Hendrickson, 1997). Similarly, an IS2-encoded promoter leads to constitutive expression from the gal operon in *E. coli*. However, activation only occurs when IS2 inserts in one orientation; in the opposing orientation IS2 represses gal expression 100-fold (Saedler et al, 1974). In *E. coli*, the cryptic bgl operon is activated by insertion of IS1 and IS5 elements, but here activation is a result of changes to cis-regulatory elements in the bgl promoter, rather than a transposon-encoded promoter (Reynolds et al, 1981; Schnetz & Rak, 1992). Gene inactivation by transposons can also be selected for rapidly in response to a strong selective pressure. Infections by *Yersinia pestis* generate a strong immune response against the capsular antigen which provides an adaptive immunity against future infections. However, mutants of *Y. pestis* readily escape this adaptive immune response as a result of IS1541 insertion into the cafA1 gene, thereby preventing assembly of the antigen without compromising virulence (Cornelius et al, 2009). Interestingly, an IS1541 insertion into the inv gene in *Y. pestis* is believed to have been a key speciation event and the loss of Inv expression from IS1541 insertion contributes to the systemic disease that makes plague so deadly (Simonet et al, 1996). IS1541 represents a case where a single transposon can contribute to evolution of a species as well as adaptation to a specific selective pressure.

In addition to causing genetic changes to the host genome, transposons themselves can directly contribute genes that provide a selective advantage to the host. By definition, these passenger genes do not contribute transposition functions but instead ensure survival and dissemination of the transposon. Passenger genes can encode determinates for virulence (Fetherston & Perry, 1994; Fetherston et al, 1992; So et al, 1979), but are best known for conferring antibiotic resistance. In this way, transposons are intimately associated with the spread of antimicrobial resistance. Penicillin and sulfonamide were the first antibiotics used clinically and resistance to both of these drugs emerged within 10 years of first use (Barber & Rozwadowska-Dowzenko, 1948; Levy & Marshall, 2004). Likewise, streptomycin-resistant *Mycobacterium tuberculosis* was identified the same year that streptomycin was introduced (Crofton & Mitchison, 1948). The number and distribution of resistant bacteria has increased dramatically, in large part because of transposon-mediated dissemination of these genes (Caniça et al, 2015).
Several transposons were first isolated from antibiotic-resistant bacteria and more transposons are discovered alongside antibiotic resistance determinates to this day (Carias et al, 1998; Ferreira et al, 2016; Petrovski et al, 2011; Rice & Carias, 1998; Warburton et al, 2007). Vancomycin and colistin are last-line antibiotics for multi-drug resistant Gram-positive and Gram-negative bacteria respectively. Genes encoding resistance to both of these antibiotics (vancomycin, \textit{vanA}; colistin, \textit{mcr-1}) have been recently found as part of composite transposons Tn1546 and ISApl1, respectively (Qureshi et al, 2014; Snesrud et al, 2016). Transposons help the spread of antibiotic resistance genes indirectly by facilitating the evolution of other mobile genetic elements (MGE) such as bacteriophage and plasmids. Transposons accomplish this by shuffling genes between MGE and chromosomes. It is also tempting to speculate that the strong selective pressure applied by antibiotic use rapidly causes the evolution of composite transposons. Here, resident IS elements can form a composite transposon with antibiotic resistance determinants and generate a new transposon which can rapidly spread antibiotic resistance genes throughout a bacterial population.

1.3 Regulation of bacterial transposons

A key component of the host-transposon relationship is that transposition must be tightly regulated. Although transposons contribute to genetic variability and often carry beneficial genes, high levels of transposition leads to genomic instability of the host (Curcio & Derbyshire, 2003). Accordingly, bacteria employ a number of mechanisms to limit the amount of transposition to protect genomic integrity. Additionally, transposons themselves exhibit self-restraint which ensures survival of the element by avoiding killing of the host.

1.3.1 Intrinsic control of transposition

The frequency of transposition in a given cell is typically determined by the amount of transposase protein present. Transposase expression is accordingly a key point of regulation for controlling transposition in bacteria.
Many endogenous promoters for transposase genes are inherently weak thus limiting transcription. For example, the promoters for IS30, IS1911, and IS10 transposase are all <10% as active as pLacUV5, while promoter activity for IS21 is undetectable in vivo (Dalrymple & Arber, 1985; Duval-Valentin et al, 2001; Reimmann et al, 1989; Simons et al, 1983). Additionally, many promoters overlap with the transposase binding site in the terminal inverted repeat of the element (Figure 1.1B, i). This allows a form of feedback inhibition where transposase expression (or in some cases a C-terminally truncated protein) inhibits its own transcription (Duval-Valentin et al, 2001; Hu et al, 1994; Zerbib et al, 1990). The production of truncated transposase proteins is also a feature common to many elements. The catalytic domain for most transposases is located in the C-terminal domain, so premature transcription termination generates an inactive transposase (Nagy & Chandler, 2004). For IS50, an internal promoter generates an N-terminally truncated transposase that forms non-productive transposase heterodimers and inhibits transposase activity (Johnson & Reznikoff, 1984).

Transposase expression is also repressed by intrinsic post-transcriptional control. Transposition into an actively transcribed gene might be expected to increase transposase expression by read-through transcription from adjacent genes. However, IS10 and IS50 encode symmetrical repeats that form an mRNA secondary structure that sequesters the translation initiation region (TIR) and blocks translation. One of these repeats is upstream of the native promoter, so only read-through transcripts form this structure (Kleckner, 1990a; Krebs & Reznikoff, 1986; Ma et al, 1994). Intrinsically weak Shine-Dalgarno sequences, programmed translational frameshifting, and leaderless mRNAs also contribute to weak translation of transposase mRNAs (Kleckner, 1990b; Nagy & Chandler, 2004). Transposons can also encode regulatory RNAs that inhibit transposase expression. IS10 encodes an antisense RNA (asRNA, RNA-OUT) that is complementary to the first 35nt of transposase mRNA (RNA-IN). Base-pairing between RNA-OUT and RNA-IN prevents ribosome binding and leads to degradation of RNA-IN (Kittle et al, 1989; Ma & Simons, 1990; Simons & Kleckner, 1983). For IS30, the asRNA (RNA-C) is complementary to the transposase coding sequence and is proposed to block translation elongation (Arini et al, 1997). The advent of transcriptomics in a range of bacteria has identified at least 17 novel asRNAs to transposases indicating that this mode of
regulation is widespread; however, only two asRNAs have been functionally characterized (reviewed in (Ellis & Haniford, 2016)).

In addition to poor expression, transposase proteins themselves are typically unstable including IS1 and IS911 transposase (Haren et al, 1997; Reif & Saedler, 1975). Lastly, formation of the transpososome (a higher order complex of transposon DNA and transposase) and the transposition reaction catalyzed by transposase are inefficient due to suboptimal protein-DNA interactions and inefficient catalysis by transposase (Gueguen et al, 2005).

1.3.2 Host factors affecting transposition

Bacterial cells control transposition by regulating almost every step in transposition including transcription initiation, translation, transposase-DNA interactions, transpososome dynamics, and post-translational control of transposase levels.

Transposition is repressed in two ways by DNA Adenine Methyltransferase (Dam). First, many transposase promoters are repressed by Dam-methylation including IS10, IS50, IS3, and IS903. Transposase promoters are almost always located at the ends of transposons and accordingly overlap with transposase binding sites. Methylation interferes with transposase-DNA interactions for IS10 and IS50 (and likely most transposases), and therefore Dam can represses expression and function of transposase proteins (Roberts et al, 1985; Spielmann-Ryser et al, 1991; Yin et al, 1988). Epigenetic control of transposons accomplishes two things: (i) transposition is linked to cell-cycle, and (ii) transposase proteins will act preferentially in cis. DNA is hemi-methylated following replication and transposase transcription will therefore be highest in the short amount of time between replication and re-establishment of methylation by Dam. Additionally, transposase will preferentially bind to hemi-methylated DNA which leads to a cis-bias where transposase proteins act preferentially on the DNA they were expressed from. There are few examples of other host factors affecting transposase expression. The global transcription factors LexA and Crp are proposed to directly control transcription of IS50 transposase (Kuan & Tessman, 1991; Ross et al, 2014). Less is known about post-transcriptional regulation of transposase expression: RNase E
degrades IS10 transposase mRNA as a consequence of translational inhibition, and the RNA-binding protein Hfq represses IS10 transposase translation (Jain & Kleckner, 1993; Ross et al, 2010).

Transposition is also controlled by the host at the post-translational level. Proteins involved in nucleoid architecture commonly control formation and stability of the transpososome (Mahillon & Chandler, 1998; Nagy & Chandler, 2004). IHF, HU, and H-NS all interact with transposon sequence in vitro and in vivo stimulate transpososome assembly or transposition (Lavoie & Chaconas, 1993; Signon & Kleckner, 1995; Surette et al, 1989; Wardle et al, 2005; Whitfield et al, 2009). The stability of transposase proteins may also be an important way that the host ensures transposase proteins act in cis. Lon protease rapidly degrades the IS903 transposase (half-life < 3 min) by recognizing the C-terminal portion of the protein. In cis, the transposase N-terminal domain can interact with transposon DNA co-translationally, resulting in the initial formation of the transpososome before the labile C-terminus is produced (Derbyshire et al, 1990). Additional host factors for transposition include Gyrase and Topisomerase I (Pato & Banerjee, 1996), DnaA (Yin & Reznikoff, 1987), and ClpX (Levchenko et al, 1995).

1.4 Transposons and stress response

In her Nobel Prize acceptance speech, McClintock proposed that the host organism could ‘activate’ transposition in response to stress (McClintock, 1984). A transient increase in transposition could lead to mutations that allow the host to adapt to extreme stress as well as ensuring the survival of the transposon. There is some evidence that transposon activity increases in response to bacterial stress. UV radiation induces IS10 transposition in E. coli (Eichenbaum & Livneh, 1998) and ISDra2 transposition in Deinococcus radiodurans (Pasternak et al, 2010), tetracycline exposure stimulates the CTnDOT conjugative transposon in Bacteroides (Shoemaker & Salyers, 1988), and IS1 insA transcription increases in E. coli in response to elevated concentrations of divalent metals (Brocklehurst & Morby, 2000). The molecular mechanisms underlying transposon
activation are not clear, although activation by host factors rather than changes in intrinsic regulation appears to be more likely. In Pseudomonas putida nutrient limitation during stationary phase growth induces the stress-responsive sigma factor, σ^S, and increases transcription of Tn4652 transposase (Ilves et al, 2001). The involvement of Hfq in repressing Tn10 and Tn5 transposition (Ross et al, 2014; Ross et al, 2010) might be another way that transposons respond to stress. Hfq is a global regulator of stress responses in bacteria and acts by promoting the action of small non-coding RNAs.

1.5 Regulatory RNA in bacteria

The central dogma of molecular biology was proposed in 1958 by Francis Crick and reinforced almost 20 years of belief that RNA acted as an intermediate between DNA and protein (Crick, 1958). This view began to change in the 1970s with the discovery of long noncoding RNAs and self-splicing catalytic RNAs in eukaryotes (Morris & Mattick, 2014). It was not until 1981 that the first regulatory RNA in bacteria was discovered. The ~108 nt RNA I inhibited replication of the ColE1 plasmid by base-pairing with the RNA that is processed to produce the replication primer (Tomizawa & Itoh, 1981). While eukaryotic regulatory RNAs continued to be discovered at an accelerating rate, bacterial regulatory RNAs were thought to be peculiar to extrachromosomal DNA such as plasmids and transposons. This view changed 15 years ago with the discovery that regulatory RNAs are widespread in bacteria and play an essential role in regulating many biological processes (Waters & Storz, 2009).

1.5.1 Base-pairing small RNAs (sRNA)

The largest class of regulatory RNA in bacteria are small non-coding RNAs (sRNA). The first sRNAs discovered in E. coli were expressed from intergenic regions of the chromosome, were 50-250 nts long, and contained a Rho-independent transcriptional terminator (Gottesman, 2004). These observations guided the first computational searches for sRNAs which together identified almost 150 novel sRNAs in E. coli, over 30 of which were experimentally validated (Argaman et al, 2001; Chen et al, 2002; Rivas et al, 2001; Wassarman et al, 2001). In recent years it has become clear that untranslated
regions of mRNAs are also a rich source of sRNAs (Chao et al, 2012; Chao & Vogel, 2016; Guo et al, 2014; Miyakoshi et al, 2015a; Miyakoshi et al, 2015b). sRNAs regulate gene expression at the post-transcriptional level by base-pairing mechanisms. Because sRNAs are expressed from loci distinct from their target, they are also referred to as trans-acting sRNA (Figure 1.2A). sRNAs share limited and often discontinuous complementarity with target mRNAs, typically consisting of 6-8 nt seed regions (perfect, continuous complementarity) with up to 20 nt of additional discontinuous complementarity.

Base-pairing between an sRNA and mRNA can increase protein synthesis but the vast majority of sRNAs repress target protein expression (common mechanisms for target repression and activation are summarized in Figure 1.2B). The translation initiation region (TIR) is a ~50-60 nt region at the 5’ end of protein coding genes that includes the Shine-Dalgarno (SD) sequence and translation initiation codon (typically AUG); the TIR begins up to 40 nts upstream of AUG and ends at the fifth translated codon (nts -40 to +15 relative to AUG) (Huttenhofer & Noller, 1994; Shine & Dalgarno, 1975). Base-pairing between an sRNA and any part of the TIR is able to inhibit translation by sterically occluding 30S ribosome binding (Bouvier et al, 2008). However, sRNAs such as GcvB and RyhB inhibit translation by base-pairing far upstream of the TIR and masking a translational enhancer and inhibiting an upstream ORF respectively (Sharma et al, 2007; Vecerek et al, 2007). Active translation typically stabilizes mRNAs by protecting from cellular ribonucleases (usually RNase E) and other components of the degradosome (Jain & Kleckner, 1993; Mohanty & Kushner, 2016). mRNA turnover is therefore a secondary effect of translational repression and ensures repression is irreversible (Morita et al, 2006). sRNAs can also directly recruit RNase E and other components of the degradosome to induce mRNA degradation and this is the main pathway for sRNAs that base-pair outside the TIR (Bandyra et al, 2012; Caron et al, 2010; Masse et al, 2003).
Figure 1.2. Common mechanisms for sRNA-mediated gene regulation.

(A) Gene schematics for bacterial sRNAs and target mRNAs. Canonical sRNAs (blue) are located in intergenic regions, are 50-250 nt long, contain a Rho-independent terminator (3’ stem-loop structure followed by a poly-U stretch), and have a promoter that is strongly induced in response to stress or external stimuli. Many sRNAs contain additional structural elements that can be important for stability and/or function. Target mRNAs (red) are encoded in a different region of the genome. The translation initiation region (TIR) is a 50-60 nt region at the 5’end of a protein coding gene and contains the ribosome binding site (encompassing the Shine-Dalgarno sequence), translation initiation codon (typically AUG), and the sequence immediately surrounding these elements. (B) Common modes of sRNA-mediated gene regulation. sRNAs (blue) base-pair with target mRNAs (red) by limited and discontinuous sequence complementarity. The consequence of base-pairing is repression (red box) or activation (green box) of protein synthesis. Most sRNAs base-pair with the TIR which blocks ribosome (grey circles) binding (i); as ribosomes typically protect mRNAs from RNase cleavage, a secondary consequence of blocking translation is degradation of the mRNA by cellular RNases (scissors). sRNA base-pairing outside of the TIR can recruit cellular RNases and induce transcript
degradation independent of inhibition of translation initiation (ii). Translation of some mRNAs is inhibited by secondary structure, and sRNAs can base-pair with an mRNA and induce structural changes that allows ribosome access (iii). Lastly, sRNAs can bind and mask RNase recognition sequences on mRNAs and prevent degradation of the mRNA (iv).
sRNAs can also increase protein expression although this mode of action is less common. In the best studied example, the 5’UTR of rpoS (the stress-responsive sigma factor, $\sigma^5$) forms a secondary structure that sequesters the TIR. DsrA binding to the rpoS 5’UTR remodels this secondary structure and activates translation (Lease et al, 1998; Majdalani et al, 1998). sRNAs can also stabilize mRNAs and therefore increase protein expression independent of controlling translation (Fröhlich et al, 2013; Papenfort et al, 2013).

Transcription of most sRNAs is induced (up to 100-fold) in response to specific stimuli such as outer membrane stress (MicA, RybB), sugar-phosphate stress (SgrS), low iron (RyhB), or oxidative stress (OxyS) (Waters & Storz, 2009). Transcriptional activation of sRNAs can occur through repressors and activators. The ferric iron uptake repressor (Fur) inhibits RyhB transcription in the presence of iron. In iron-limiting conditions, RyhB expression is de-repressed and RyhB then inhibits expression of iron-storage and -binding proteins (e.g. *sdhCDAB*) to increase iron availability in the cell (Masse et al, 2003; Massé & Gottesman, 2002). In contrast, sugar-phosphate stress (e.g. accumulation of glucose-6-phosphate) leads to activation of SgrR which is a transcriptional activator of SgrS (Vanderpool & Gottesman, 2004). SgrS combats sugar-phosphate stress by repressing glucose transporters and stabilizing the mRNA for a sugar phosphatase (Kawamoto et al, 2006; Papenfort et al, 2013; Rice et al, 2012). However, some sRNAs are expressed and repress their targets constitutively. ChiX is transcribed constitutively in the absence of chitin and represses the chitin transporter *chiP*. The presence of chitin induces degradation of ChiX, thereby increasing *chiP* expression (Figueroa-Bossi et al, 2009; Rasmussen et al, 2009). sRNAs can be synthesized faster than protein regulators and can therefore provide a rapid response, in many cases acting within minutes of induction to regulate target mRNAs. sRNAs also have an advantage over protein regulators because the regulatory molecule does not need to be translated, which conserves time and resources within the cell. This ensures bacteria are able rapidly and efficiently adapt to stress conditions. Finally, sRNAs can target multiple mRNAs which allows the coordinate expression of a network of functionally related proteins (Wagner & Romby, 2015).
1.5.2 Antisense RNAs (asRNA)

Antisense RNAs (asRNAs) are often considered a subclass of sRNAs, with the key difference being that they are encoded on the opposite strand of DNA to the mRNA they regulate (also called cis-acting sRNA). Accordingly, asRNAs share extensive and perfect complementarity (30 to >300 nt) with their mRNA (Wagner et al, 2002). Most asRNAs have been identified in MGE (plasmids, transposon, and bacteriophage) although transcriptomics experiments have identified asRNAs as a natural component of many bacterial genomes; over 300 putative asRNAs were discovered in *E. coli*, 21 of which have been experimentally validated (Lybecker et al, 2014; Sesto et al, 2013). Like sRNAs, asRNAs act by base-pairing with mRNAs although the consequences of pairing are diverse (Figure 1.3). asRNAs can regulate transcription attenuators, stabilize or induce degradation of mRNAs, and repress translation (Brantl, 2007). Unlike sRNAs, most asRNAs are constitutively expressed. Because of extensive complementarity between asRNAs and mRNAs, pairing results in an extended duplex that is frequently targeted for degradation by RNase III. A single asRNA can also regulate divergently encoded genes. Here, an asRNA for one gene/operon can increase transcription of the neighboring gene/operon. The result is that a single asRNA represses one group of genes while activating a neighboring set of genes, and this arrangement is accordingly referred to as an excludon (Sesto et al, 2013). In most cases the asRNA ensures discordant expression of genes with opposing functions.
Antisense RNAs (asRNA) regulate (A) transcription, (B) translation, (C,D) plasmid replication, and (E,F) can affect mRNA stability. Antisense RNAs are shown in red, mRNAs in blue, and black rectangles indicate promoters. Black and green arrows indicate the action of RNases. Reprinted from Current Opinion in Microbiology, Volume 10, S. Brantl, Regulatory mechanisms employed by cis-encoded antisense RNAs, Pages 102-109, 2007, with permission from Elsevier.
1.5.3 RNAs that regulate protein activity and riboswitches

Regulatory RNAs can also act by antagonizing the action of RNA-binding proteins. CsrA (carbon storage regulator) regulates gene expression at the post-transcriptional level by binding GGA motifs typically located in the 5’UTR of target mRNAs (Babitzke & Romeo, 2007). CsrA binding to mRNAs usually inhibits translation by blocking ribosome access and translation inhibition leads to transcript turnover. CsrA binding can also stabilize mRNAs by an unclear mechanism. Two regulatory RNAs, CsrB and CsrC, antagonize CsrA by mimicking substrates and titrating CsrA away from target mRNAs. Similarly, the 6S RNA mimics an open promoter and sequesters σ^70 containing RNA polymerase ultimately repressing transcription of σ^70-dependent promoters (Wassarman, 2007).

Riboswitches are structural elements in mRNAs that undergo conformation changes to regulate transcription or translation. These cis-regulatory elements differ from the above examples of regulatory RNA because they are not expressed as an independent RNA, instead they regulate the same molecule of RNA they are encoded on. Riboswitches can sense stalled ribosomes, changes in temperature (also referred to as RNA thermometers), tRNA availability, or even small molecule ligands (Grundy & Henkin, 2006; Waters & Storz, 2009).

1.6 Hfq

Hfq was first identified in 1968 as a host factor for Qβ phage replication (Franze de Fernandez et al, 1968). Early studies of Hfq revealed that it is an abundant, heat-stable protein that binds single-stranded AU-rich sequences on RNA (de Haseth & Uhlenbeck, 1980a; de Haseth & Uhlenbeck, 1980b; Franze de Fernandez et al, 1972; Hori & Yanazaki, 1974; Senear & Steitz, 1976). The first hint of the importance of Hfq was the finding that a hfq mutant was extremely pleiotropic, exhibiting decreased growth rate and altered cell morphology, increased sensitivity to osmotic and UV stress, and decreased plasmid supercoiling (Tsui et al, 1994). Hfq is now recognized as a central component of
many bacterial stress and virulence responses, acting as a chaperone for sRNAs (Papenfort & Vogel, 2010; Vogel & Luisi, 2011).

1.6.1 Structure and RNA-binding properties

Hfq is a member of the Sm/LSm family of RNA-binding proteins and is present in over half of all sequenced bacteria (Sobrero & Valverde, 2012; Sun et al, 2002). The E. coli Hfq monomer is 102 amino acids (11.2 kDa) and consists of a compact LSm domain (consisting of an Sm1 and Sm2 motif) with a ~40 residue intrinsically disordered C-domain (CTD). Hfq monomers assemble into a toroidal homohexamer that contains three RNA-binding surfaces (Figure 1.4; reviewed in (Brennan & Link, 2007; Sauer, 2013; Updegrove et al, 2016)).

The N-terminus of each monomer forms the proximal binding surface (located on the ‘top’ of the hexamer). A nucleotide binding pocket located in the monomer-monomer interface interacts primarily with uracil (U; to a lesser extent adenine, A) residues and accordingly the proximal surface can interact with a U₆ motif. A group of glutamine (Q8) and lysine residues (K41, K56, K57) together provide specificity for U and A and a phenylalanine (F42) is involved in base-stacking interactions. The proximal surface is thought of as the sRNA-binding surface and preferentially interacts with the poly-U tail following the Rho-Independent terminator found on almost all sRNAs. Additional specificity for Rho-Independent terminators comes from preferential binding to structured regions and 3’OH (Ishikawa et al, 2012; Sauer & Weichenrieder, 2011).

The opposing distal surface of Hfq (located on the ‘bottom’ of the hexamer) is viewed as the mRNA-binding surface. Each monomer of Hfq binds an ARN triplet, where R refers to either purine and N is any nucleotide. Discrimination for the A site is largely due to hydrogen bonding between glutamine (Q33) and N₆/₇ on adenine, and a conserved tyrosine (Y25) forms base-stacking interactions with the purine in the R site (Link et al, 2009). The preferred binding motif for the distal surface on the hexamer is accordingly (ARN)₆.
Figure 1.4. Structure of Hfq.

(A) Secondary structural elements highlighting the conserved Sm1 and Sm2 motifs in Hfq. (B,C) Each protomer is a $\alpha$-$\beta_{1-5}$ structural unit (i.e. one $\alpha$-helix and five $\beta$-strands) in which the $\beta$-strands form antiparallel sheets. The $\beta_2$-strand is twisted and curved to such an extent that it contributes to both sheets to form a self-closing, squat barrel. The N-terminal helix and squat barrel are structural signatures of Hfq-Sm-LSm proteins. The $\beta_4$ and $\beta_5$ strands on the periphery of each Hfq subunit expose hydrogen-bonding edges that interact with the strands of the neighboring protomers, so that sheets effectively continue over the entire ring. (D) Two faces for interaction with RNAs (orange) are presented on opposite sides of the Hfq ring. The proximal face (the surface on which the N-terminal $\alpha$-helix is exposed) includes residues in the Sm2 sequence motif. Disordered tails are likely to emanate from the equator of the Hfq ring and may form distributive electrostatic interactions with nucleic acids. This figure and legend is adapted by permission from Macmillan Publishers Ltd: NATURE REVIEWS MICROBIOLOGY (Vogel & Luisi, 2011), copyright 2011.
The lateral (or rim) surface of Hfq connects the proximal and distal surfaces. In Gram-negative bacteria, this surface is extremely basic and is mostly thought to interact non-specifically with RNA. A cluster of arginine residues (R17/18/19) are required for Hfq’s chaperone activity in many systems, and this surface is proposed to be the active site for sRNA-mRNA pairing (see below) (Panja et al, 2013). Intriguingly, the lateral surface is acidic in gram-positive bacteria where Hfq is mostly uninvolved in post-transcriptional regulation (Sobrero & Valverde, 2012; Vogel & Luisi, 2011; Zheng et al, 2016).

The role of the Hfq CTD has remained controversial (Sobrero & Valverde, 2012; Updegrove et al, 2016) but recent work indicates that the CTD plays an important role in promoting the release of sRNA-mRNA paired species (see below) and regenerating Hfq for further catalysis (Santiago-Frangos et al, 2016). The CTD is proposed to ‘sweep’ the proximal surface and promote release of sRNAs thereby preventing the sequestration of Hfq in nonproductive Hfq-sRNA complexes.

**1.6.2 Role of Hfq in post-transcriptional regulation**

The canonical role of Hfq in post-transcriptional regulation is to facilitate sRNA-mediated regulation of mRNAs (Hopkins et al, 2011; Moller et al, 2002; Soper & Woodson, 2008; Zhang et al, 2002). Hfq is believed to accomplish this with dedicated RNA-binding surfaces: the proximal surface binds sRNAs, the distal surface binds mRNAs, and the rim contributes to non-specific RNA interactions (Figure 1.5A). First and foremost, Hfq stabilizes almost all sRNAs by protecting from RNases and the degradosome (Figure 1.5B). For example, the half-life of RyhB decreases from >30 min in a WT strain to <1 min in an hfq-null strain (Masse et al, 2003). Second, Hfq is required for the function of almost all sRNAs in vivo and accelerates the rate of sRNA-mRNA pairing in vitro. The mechanism for Hfq-catalyzed pairing appears to vary based on the specific sRNA-mRNA pair (Schu et al, 2015; Zhang et al, 2013). In the simplest model, Hfq binds both an sRNA and mRNA pair and increases the local concentration of each RNA as well as correctly orienting the pairing sequences on each RNA (Figure 1.5C, i) (Panja & Woodson, 2012). The positively charged lateral surface may also provide an electrostatically favorable environment for nucleating pairing (ii) (Panja et al, 2013; Zheng et al, 2016).
Figure 1.5. Main functions of Hfq in post-transcriptional regulation.

(A) Cartoon of Hfq hexamers highlighting the three RNA-binding surface. The proximal surface (blue) interacts with sRNAs and is located on the ‘top’ of the toroidal homohexamer. The distal surface (red) is on the opposing surface and interacts with mRNAs. The lateral surface or rim (purple) is on the side of the hexamer and connects the distal and proximal surfaces. The unstructured CTD of each monomer is shown with black lines and is omitted for clarity in the rest of the figure. (B) Hfq binds the Rho-independent terminator at the 3’ end of sRNAs (blue) to protect from degradation by cellular RNases (scissors). (C) The proximal surface of Hfq binds an sRNA (blue) and the cognate mRNA (red) interacts with the distal surface (on the same hexamer). Pairing between complementary sequences (yellow) can be accelerated by Hfq increasing the local concentration and/or orienting the RNAs correctly (i). Alternatively, pairing is
catalyzed on the lateral surface (ii). (D) Pairing sequences on sRNAs (blue) or mRNAs (not shown) can be sequestered in secondary structure. Hfq binding can remodel the RNA to expose this sequence.
Lastly, Hfq can actively remodel an sRNA and/or mRNA to expose pairing sequences (Figure 1.5D) (Geissmann & Touati, 2004; Soper et al, 2011; Updegrove et al, 2016). The requirement of Hfq in sRNA-mRNA pairing partially stems from the instability of an sRNA-mRNA duplex. For this reason Hfq was not believed to be important for asRNAs, which share extensive complementarity with their partner mRNAs (Gottesman & Storz, 2011; Thomason & Storz, 2010). In vitro experiments have demonstrated that sRNA-mRNA complexes formed in the presence of Hfq are more stable than those formed without (Soper et al, 2010; Soper & Woodson, 2008), and it’s believed that Hfq promotes more extensive pairing than could occur in the absence of Hfq.

With regards to facilitating pairing, Hfq is an accessory protein for post-transcriptional control while sRNAs are the effector molecule. There is some evidence that Hfq can act autonomously to repress translation (Salvail et al, 2013; Vecerek et al, 2005; Vytvytska et al, 2000) as well as one example where the role of Hfq and an sRNA are switched. The sRNA Spot42 recruits Hfq to the TIR of sdhC mRNA and the association of Hfq with sdhC represses translation (Desnoyers & Masse, 2012). In this case Hfq is the effector protein and Spot42 is an accessory factor for translational control.

1.7 Mobile genetic elements and pathogenesis of _Salmonella Typhimurium_

Many bacterial pathogens contain foreign DNA that has been acquired over evolutionary time and contributes to virulence (Frost et al, 2005). For the closely related organisms, _E. coli_ and _Salmonella enterica_ subsp. _enterica_ serovar Typhimurium, all phenotypic differences are due to horizontally acquired genes (Ochman et al, 2000). In _S. Typhimurium_, pathogenesis is almost exclusively due to genes encoded in prophage (bacteriophage lysogen) and large segments (10-200 kb) of horizontally acquired DNA referred to as _Salmonella_ pathogenicity islands (SPI) (Figueroa-Bossi et al, 2001; Gyles & Boerlin, 2014; Juhas et al, 2009; Moreno Switt et al, 2012).
S. Typhimurium is a foodborne pathogen that causes a generally self-limiting gastrointestinal illness and is the main cause of hospitalizations and deaths due to foodborne disease in the United States (Elaine et al, 2011). S. Typhimurium and related serovars of S. enterica cause approximately 155,000 deaths worldwide, with over 93 million cases each year (Feasey et al). S. Typhimurium employs two main type-III secretion systems (T3SS) encoded by SPIs (SPI-1 and SPI-2) to infect mammalian hosts. T3SS have two components: (i) structural proteins that assemble to form the needle complex in the bacterial membrane, and (ii) effector proteins that are injected into the cytosol of host-cells (de Jong et al, 2012). The two main T3SS used by S. Typhimurium are encoded by SPI-1 and SPI-2.

SPI-1 is a ~40kb genomic island that encodes the T3SS required for invasion of non-phagocytic cells. SPI-1 is crucial for S. Typhimurium crossing the intestinal epithelium in infected hosts (Lostroh & Lee, 2001; Que et al, 2013). SPI-1 encoded effector proteins are injected into the host cell and perturb host signal transduction pathways to induce cytoskeleton rearrangements and membrane ruffling, leading to the engulfment of the bacteria in a phagosome (Salmonella containing vacuole, SCV). SPI-2 genes become activated inside the SCV and are important for intracellular survival (Ibarra & Steele-Mortimer, 2009).

1.8 Scope of this thesis

There is a wealth of information about transcriptional and post-translational regulation of bacterial transposons, but much less is known about post-transcriptional regulation of transposase expression. In Chapters 2 and 3 I investigate post-transcriptional regulation of the model transposon, Tn10/IS10, and show that the host bacterium regulates transposition at the post-transcriptional level. I provide evidence in Chapter 2 that the E. coli RNA-binding protein Hfq catalyzes antisense pairing between IS10 transposase mRNA (RNA-IN), and a cis-encoded asRNA (RNA-OUT). Hfq was previously implicated in translational control of IS10 transposase expression (Ross et al, 2010), but exactly how Hfq accomplishes this was unclear. My work shows that Hfq alters the
secondary structure of both RNA-IN and RNA-OUT to expose sequence otherwise sequestered in secondary structure. I present work in Chapter 3 that shows that Hfq regulates IS10 transposase expression in an antisense-independent manner. In the absence of RNA-OUT, Hfq represses transposase translation by directly binding to the TIR and blocking ribosome binding. This is one of only a few examples where Hfq regulates protein synthesis in the absence of a regulatory RNA. Additionally, I show that induction of an sRNA can perturb transposon regulation by Hfq and this provides evidence that Hfq links bacterial stress responses to transposition at the post-transcriptional level.

In Chapter 4, I expand my studies to a unrelated transposon, IS200. The identification of an asRNA encoded by IS200 led us to ask if Hfq regulates a second antisense system in a similar way to IS10. Unlike IS10, Hfq does not participate in antisense regulation but does directly repress translation. I show Hfq works synergistically with an asRNA and mRNA secondary structure in the transposase mRNA to repress transposition. Importantly, we demonstrated for the first time that IS200 transposition could be measured under standard laboratory conditions, but only when we disrupted post-transcriptional control. My work provides an explanation for the long-standing observation that IS200 elements are essentially dormant transposons. This work also shows that the host bacteria and transposon encode negative regulators of transposition that work together to inhibit a single step in transposition (transposase translation initiation).

Lastly, I present evidence in Chapter 5 that a transposon-encoded RNA regulates gene expression in S. Typhimurium. The finding that IS200 is silenced by three seemingly redundant post-transcriptional mechanisms prompted me to ask if IS200-encoded RNAs serve a transposition-independent role in S. Typhimurium. IS200 elements are highly conserved and are present in multiple copies in many Salmonella enterica genomes, despite being inactive transposons. I speculated that there might be a selective advantage to the host bacteria by maintaining IS200 elements in high copy-number. Since the transposase protein is almost never synthesized, I profiled gene expression in S. Typhimurium under conditions where expression of IS200-encoded
RNAs was altered. My work shows that the IS200 transposase mRNA is processed to produce a trans-acting sRNA that regulates SPI-1 gene expression in S. Typhimurium.

1.9 References


Barth PT, Datta N, Hedges RW, Grinter NJ (1976) Transposition of a deoxyribonucleic acid sequence encoding trimethoprim and streptomycin resistances from R483 to other replicons. *Journal of bacteriology* **125**: 800-810


Cariñas LL, Rudin SD, Donskey CJ, Rice LB (1998) Genetic linkage and cotransfer of a novel, vanB-containing transposon (Tn5382) and a low-affinity penicillin-binding protein 5 gene in a clinical vancomycin-resistant Enterococcus faecium isolate. *Journal of bacteriology* **180**: 4426-4434


Fetherston JD, Perry RD (1994) The pigmentation locus of Yersinia pestis KIM6+ is flanked by an insertion sequence and includes the structural genes for pesticin sensitivity and HMWP2. *Molecular microbiology* **13**: 697-708
Fetherston JD, Schuetze P, Perry RD (1992) Loss of the pigmentation phenotype in Yersinia pestis is due to the spontaneous deletion of 102 kb of chromosomal DNA which is flanked by a repetitive element. *Molecular microbiology* **6**: 2693-2704


Kittle JD, Simons RW, Lee J, Kleckner N (1989) Insertion sequence IS10 anti-sense pairing initiates by an interaction between the 5' end of the target RNA and a loop in the anti-sense RNA. *Journal of molecular biology* **210**: 561-572

Kleckner N (1990a) Regulating tn10 and is10 transposition. *Genetics* **124**: 449-454


Lee H, Doak TG, Popodi E, Foster PL, Tang H (2016) Insertion sequence-caused large-scale rearrangements in the genome of Escherichia coli. *Nucleic acids research*


Petrovski S, Blackmore DW, Jackson KL, Stanisich VA (2011) Mercury(II)-resistance transposons Tn502 and Tn512, from Pseudomonas clinical strains, are structurally different members of the Tn5053 family. *Plasmid* **65**: 58-64


Reif HJ, Saedler H (1975) IS1 is involved in deletion formation in the gal region of E. coli K12. *Molecular & general genetics : MGG* **137**: 17-28


Ross JA, Trussler RS, Black MD, McLellan CR, Haniford DB (2014) Tn5 transposition in Escherichia coli is repressed by Hfq and activated by over-expression of the small non-coding RNA SgrS. *Mob DNA* **5**: 27


Siguier P, Gourbeyre E, Chandler M (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS microbiology reviews* **38**: 865-891


Surette MG, Lavoie BD, Chaconas G (1989) Action at a distance in Mu DNA transposition: an enhancer-like element is the site of action of supercoiling relief activity by integration host factor (IHF). The EMBO journal 8: 3483-3489

Taylor AL (1963) BACTERIOPHAGE-INDUCED MUTATION IN ESCHERICHIA COLI. Proceedings of the National Academy of Sciences of the United States of America 50: 1043-1051


Chapter 2

2 Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the Tn10/IS10 system

2.1 Introduction

Small RNAs (sRNAs) have emerged as important components of gene expression regulatory networks in bacteria. sRNAs generally function by base-pairing to mRNAs with which they share at least partial sequence complementarity. Base-pairing between sRNAs and mRNAs typically influences translation and/or stability of the mRNA (for reviews, see (Gottesman & Storz, 2011; Vogel & Luisi, 2011)). sRNAs are categorized as either trans- or cis-encoded. Trans-encoded sRNAs are expressed from distinct loci relative to the transcripts they regulate, whereas cis-encoded sRNAs are expressed from the strand opposite their target mRNA and, consequently, are perfectly complementary to at least a portion of their target RNA (Figure 2.1A). Cis-encoded sRNAs are also referred to as antisense RNAs (or asRNA). The regulation imposed by many trans-encoded sRNAs is dependent on the protein Hfq, an Sm-family protein that is present in many bacterial species. With regard to sRNA-based regulation, Hfq functions by promoting the pairing of sRNAs to their target mRNAs (Moller et al, 2002; Vogel & Luisi, 2011; Zhang et al, 2002). The expression of many trans-encoded sRNAs is up-regulated by environmental stress, and typically this imposes a biological response to stress through Hfq-mediated pairing of sRNAs and their target mRNAs (Altuvia et al, 1997; Vogel & Papenfort, 2006)

asRNAs were originally found on extrachromosomal DNAs, such as plasmids and transposons. Chromosomally encoded asRNAs have since been identified, although a subset of these is imbedded within mobile DNA elements that have recently been acquired (e.g., pathogenicity islands). The general perception is that, due to the perfect

---

(A) Cis- vs. trans-encoded sRNAs. Transcribed strands of three different genes and their corresponding RNAs (colour coded) are shown. Pairing of a trans-sRNA (gold) and an mRNA (green) and of a cis-sRNA (pink) and an mRNA (cyan) is shown. Hfq (blue hexamer) catalyzes pairing in the former case where there is partial sequence complementarity between partners, but it is unclear if it also catalyzes pairing in the latter case when there is perfect sequence complementarity between partners. Asterisks (*) define the translation initiation region (TIR) of the mRNAs. (B) Structure of Tn10 and IS10-Kan. Tn10 is a 9147 bp composite transposon that confers tetracycline resistance (Tet\(^R\)). Tn10 is comprised of IS10-left and IS10-Right, the latter of which encodes a functional transposase protein that catalyzes DNA cleavage and joining events involving the “outside” (OE) and “inside” (IE) ends. The transposase mRNA (RNA-IN) is encoded from the promoter plN (blue squares). A second promoter (pOUT – black squares) within IS10-Right encodes a cis-sRNA (also referred to as an antisense RNA), RNA-OUT. To follow transposition of IS10-Right in *E. coli*, a Kan\(^R\) gene cassette was cloned into IS10-Right, creating IS10-Kan. RNA-OUT is depicted as a stable stem-loop structure (black) and RNA-IN is depicted as a blue line with asterisks defining the TIR. RNA-OUT is known to pair with RNA-IN, and this inhibits translation of RNA-IN, thereby down-regulating transposition. Hfq can enhance the rate of RNA-IN:OUT pairing *in vitro*, but it is not known if Hfq plays a role in this antisense system *in vivo*. 

Figure 2.1. Small regulatory RNAs (sRNAs) and the Tn10/IS10 antisense system.
complementarity between asRNAs and their targets, the regulatory function of asRNAs will not be dependent on Hfq (Waters & Storz, 2009).

Tn10/IS10 (Figure 2.1B) encodes a 69 nt asRNA (RNA-OUT) that regulates transposase expression by pairing with the transposase mRNA (RNA-IN). This pairing down-regulates transposase translation by sequestering the 5’ translational initiation region (TIR) from the ribosome, thereby limiting transposase translation (Ma & Simons, 1990; Simons & Kleckner, 1983). We recently demonstrated that the frequency of Tn10/IS10 transposition from a multi-copy plasmid is greatly increased in an hfq− strain of Escherichia coli, thereby implicating Hfq as a potent negative regulator of Tn10/IS10 transposition (Ross et al., 2010). The Hfq-effect was much less robust in a system with reduced RNA-OUT levels, suggesting that Hfq functions, in part, through antisense regulation. Transposase expression from a translational fusion was also found to increase in the hfq− background, consistent with a post-transcriptional role for Hfq in transposase regulation. Furthermore, studies in vitro demonstrated that Hfq bound both RNA-IN and RNA-OUT and increased the rate at which these molecules pair. Taken together, these results are consistent with Hfq playing an important role in translational regulation mediated by an asRNA (Ross et al., 2010).

The finding that Hfq participates in a regulatory system involving an asRNA raises questions regarding the mechanism through which Hfq acts in this system and invites comparisons to its mechanism of action in trans-sRNA regulated systems. Hfq readily forms a hexamer, and the hexameric unit possesses at least two RNA binding sites. These sites, referred to as the proximal and distal binding sites, are located on opposing surfaces of the toroidal structure of the hexamer. Structure-based design of mutations in the proximal and distal binding sites has been important in defining the RNA binding specificities of these sites. The proximal site mutation K56A blocks Hfq binding to U-rich trans-sRNAs, thereby implicating the proximal site as the trans-sRNA binding site. In contrast, the Y25A mutation impairs Hfq binding to A-rich RNAs as well as mRNAs, thereby implicating the distal site in mRNA binding (Brennan & Link, 2007; Mikulecky et al., 2004; Olejniczak, 2011).
In the current work, we have further evaluated the interactions between \textit{E. coli} Hfq and Tn10/IS10-encoded RNA-IN and RNA-OUT and have begun to evaluate the importance of these interactions with respect to the function of the asRNA system of this transposon.

### 2.2 Results

#### 2.2.1 Hfq binds RNA-IN approximately 80-fold more tightly than RNA-OUT

In previous work, we demonstrated using Electrophoretic Mobility Shift Assay (EMSA) that \textit{E. coli} Hfq binds RNA-IN and RNA-OUT (Ross et al, 2010). Our initial objective in the current work was to quantify the binding strength of these interactions to facilitate comparison with previously defined trans-sRNA/mRNA partners whose pairing is catalyzed by Hfq. Towards this end, we prepared $^{32}$P-labeled RNA-OUT (69 nt + 2 extra nucleotides encoded by the expression construct) and a truncated form of RNA-IN (the first 160 nt) by \textit{in vitro} transcription and individually mixed each of these RNAs (~0.1 nM) with purified Hfq over a broad range of Hfq concentrations. Binding reactions were then analyzed on a 6% polyacrylamide gel.

We show in Figure 2.2A that two distinct Hfq-bound RNA-OUT complexes were generated in our ‘Hfq titration’. Just under 50\% of the input RNA was shifted to a reduced mobility (Hfq:OUT-1) at an Hfq concentration of 14 nM (lane 3). At 38 nM Hfq (lane 9) a second complex (Hfq:OUT-2) was detected and at 48 nM Hfq:OUT-2 and Hfq:OUT-1 were present at close to a 1:1 ratio. Apparent dissociation constants $K_{D1}$ and $K_{D2}$, for Hfq:OUT-1 and Hfq:OUT-2 respectively, are 19.6 and 44.8 nM, calculated per Hfq hexamer (Figure 2.2B and Table 2.1).

We show in Figure 2.2C that Hfq also formed multiple complexes (4 distinct species) with RNA-IN-160. Hfq:IN-1 formed at the lowest Hfq concentration in the titration. The apparent $K_D$ for this complex is $\sim$ 0.24 nM per hexamer (Figure 2.2D and
Table 2.1). This represents an 81-fold higher affinity relative to Hfq binding to RNA-OUT. At higher Hfq concentrations, additional Hfq:IN complexes were formed (Hfq:IN-

![A](image1.png) ![B](image2.png) ![C](image3.png) ![D](image4.png)

**Figure 2.2.** Hfq binds with high and moderate affinities to RNA-IN and RNA-OUT *in vitro.*

$^{32}$P-labeled RNA-OUT (A) or RNA-IN (C) was mixed with varying concentrations (reported per hexamer) of purified Hfq protein, and the reactions were subject to EMSA as described in Materials and Methods. Band intensities were quantified (ImageQuant), and the percent of each shifted species (relative to total labeled RNA) was plotted vs Hfq concentration (B, D). RNA-OUT formed two complexes with Hfq, Hfq:OUT*1, and Hfq:OUT*2. RNA-IN formed four complexes with Hfq, Hfq:IN*1, Hfq:IN*2, Hfq:IN*3, and Hfq:IN*4. Apparent dissociation constants ($K_D$) are indicated. See Table 2.1 for summary of $K_D$ values and Hill coefficients determined in this study. RNA-OUT and RNA-IN were present at a final concentration of ~0.1 nM. Error bars represent standard error from two experiments. $K_D$ is reported ± standard error.
Table 2.1. *In vitro* binding of RNA-OUT or RNA-IN to WT, distal-, or proximal-impaired Hfq mutants.

<table>
<thead>
<tr>
<th>RNA Species</th>
<th>Hfq Variant</th>
<th>$K_{D1}$ (nM)</th>
<th>$K_{D2}$ (nM)</th>
<th>h</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA-OUT</td>
<td>Hfq&lt;sub&gt;WT&lt;/sub&gt;</td>
<td>19.6 ± 0.94</td>
<td>44.8 ± 2.57</td>
<td>3.45 ± 0.49</td>
</tr>
<tr>
<td></td>
<td>his&lt;sub&gt;6&lt;/sub&gt;-Hfq&lt;sub&gt;WT&lt;/sub&gt;</td>
<td>75.6 ± 6.98</td>
<td>179 ± 18.5</td>
<td>2.20 ± 0.36</td>
</tr>
<tr>
<td></td>
<td>his&lt;sub&gt;6&lt;/sub&gt;-Hfq&lt;sub&gt;Y25A&lt;/sub&gt;</td>
<td>94.3 ± 6.50</td>
<td>202 ± 2.50</td>
<td>2.42 ± 0.32</td>
</tr>
<tr>
<td></td>
<td>his&lt;sub&gt;6&lt;/sub&gt;-Hfq&lt;sub&gt;K56A&lt;/sub&gt;</td>
<td>-</td>
<td>389 ± 14.1</td>
<td>-</td>
</tr>
<tr>
<td>RNA-IN</td>
<td>Hfq&lt;sub&gt;WT&lt;/sub&gt;</td>
<td>0.24 ± 0.01</td>
<td>1.18 ± 0.12</td>
<td>2.44 ± 0.27</td>
</tr>
<tr>
<td></td>
<td>his&lt;sub&gt;6&lt;/sub&gt;-Hfq&lt;sub&gt;WT&lt;/sub&gt;</td>
<td>0.99 ± 0.13</td>
<td>2.29 ± 0.31</td>
<td>2.09 ± 0.55</td>
</tr>
<tr>
<td></td>
<td>his&lt;sub&gt;6&lt;/sub&gt;-Hfq&lt;sub&gt;Y25A&lt;/sub&gt;</td>
<td>1.69 ± 0.23</td>
<td>11.1 ± 2.92</td>
<td>1.32 ± 0.20</td>
</tr>
<tr>
<td></td>
<td>his&lt;sub&gt;6&lt;/sub&gt;-Hfq&lt;sub&gt;K56A&lt;/sub&gt;</td>
<td>3.10 ± 0.40</td>
<td>15.1 ± 1.01</td>
<td>1.00 ± 0.10</td>
</tr>
</tbody>
</table>

Binding observed by EMSA. The percentages of RNA bound by Hfq were plotted vs Hfq<sub>6</sub> concentration (in nM), and the data were fit to a binding curve to determine apparent $K_D$ values and Hill slopes (h), expressed ± standard error. Binding assays were performed twice.
2; Hfq:IN-3 and Hfq:IN-4). It appears as though Hfq:IN-2 was generated from Hfq:IN-1, Hfq:IN-3 was generated from Hfq:IN-2 and Hfq:IN-4 was generated from Hfq:IN-3, as the appearance of each of these species coincided with the reduction in the amount of the species with the next highest gel mobility. The apparent $K_D$ and Hill Coefficient for each Hfq:RNA complex is summarized in Table 2.1.

Super-shifting in the above experiments can most easily be explained by each of the RNAs having multiple Hfq binding sites with different affinities. Detection of super-shifting over a narrower Hfq concentration range for RNA-OUT versus RNA-IN (reflected in the higher Hill Coefficient) is consistent with a higher degree of cooperativity in the former.

2.2.2 RNase and hydroxyl radical footprinting of RNA-IN, RNA-OUT, and Hfq:RNA complexes.

To further characterize Hfq:RNA-IN and Hfq:RNA-OUT interactions in vitro we used a combination of hydroxyl radical and ribonuclease (RNase) footprinting. Hydroxyl radical and RNase footprinting have both been used to identify Hfq binding sites within target mRNA and sRNAs (Brescia et al, 2003; Lease & Woodson, 2004; Rolle et al, 2006; Vecerek et al, 2005). RNase footprinting also provides insight into the structure of the RNA as well as structural changes in the RNA upon protein binding. Structure-probing techniques have not previously been applied to RNA-IN and RNA-OUT, although a model for RNA-OUT (Model I) was proposed based on predictions from in silico RNA folding programs and genetic data (see Figure 2.3B) (Case et al, 1989; Kittle et al, 1989).

For RNase structure probing/footprinting, 5’ end-labeled RNA-OUT was treated with either RNase A, T1 or V1 in the presence or absence of purified Hfq protein. RNase A and T1 cleave RNA following single stranded C/U and G, respectively, while RNase V1 cleaves 3’ of paired nucleotides.
Figure 2.3. Structure-probe analysis of RNA-OUT and Hfq:RNA-OUT complex.

(A) $^{32}$P-labeled RNA-OUT (65 nM) was incubated with or without Hfq as indicated before hydroxyl radical (lanes 4-7) or ribonuclease (A, T1, or V1; lanes 8-19) treatments. Reactions, including untreated RNA (lanes 2,3) and a G-ladder (lane 1), were analyzed on a 10% denaturing polyacrylamide gel. Nucleotide labeling is relative to the RNA-OUT in vitro transcriptional start site, which includes two extra nucleotides introduced by T7 RNA polymerase at the 5’end of the RNA. Where Hfq was included, it was present at 1460, 2190, and 4380 nM. (B) A previous model of RNA-OUT (Model I) is compared to the model derived from the current work (Model II). Colored letters represent RNase-sensitive positions in RNA-OUT observed in the absence of Hfq. Red indicates cleavage by either RNase A or T1, while blue indicates cleavage by RNase V1. Symbols (triangles and asterisks) are defined in the text.
RNase probing of RNA-OUT yielded a predicted structure similar to what has been previously proposed (Figure 2.3B). However, our data support some modifications to this model (see Figure 2.3A and B; note that ribonuclease-sensitive residues in Model II are colored red for A or T1 and blue for V1). U33, which was previously predicted to be in the unpaired loop, exhibited moderate sensitivity to V1 and relatively low sensitivity to A, suggesting that this residue is base-paired, presumably to A39. C42 and C43 were both strongly sensitive to A and this is consistent with a two-nucleotide bulge (bulge 1) in the 3’ side of the stem, immediately adjacent to the loop. Also, in Model II bulge 2 is larger than in Model I, consisting of 6 as opposed to 3 unpaired residues. This is supported by sensitivity of U24, A25, U26 and U27 to A and low sensitivity of G48 to T1. Additionally, for 14 of the 23 base-pairs in the predicted stem, at least one nucleotide in the pair showed sensitivity to V1 and only 2 base-pairs had one nucleotide that was sensitive to A. The V1-sensitive residues appeared in all 4 segments of the stem that are separated by bulges. In contrast, no residues in the loop or bulges exhibited V1 sensitivity.

Addition of Hfq to RNA-OUT caused some significant changes in the RNase cleavage profile (indicated by upward and downward pointing triangles in Figure 2.3A and B). In this experiment Hfq was added to RNA-OUT at concentrations sufficient (based on EMSA data; Figure S2.1A) to give greater than 90% Hfq:OUT-1 (1460 nM) and greater than 90% Hfq:OUT-2 (2190 and 4380 nM). Based on similarities of sample loading and total reactivity relative to the ‘no Hfq’ control, results for the ‘intermediate’ Hfq concentration (2190 nM – lanes 10, 14 and 18) were the easiest to analyze and are discussed in detail below.

Hfq binding appears to destabilize the base-paired stem of RNA-OUT. All four of the stem regions contained residues that increased in sensitivity to single strand-specific ribonuclease (upward facing red triangles) and the lower stem also contained residues that exhibited reduced sensitivity to double strand-specific ribonuclease (downward facing blue triangles). Destabilization of the stem by Hfq could be functionally significant because the 5’ portion of the stem, including residues 6-33, is expected to pair with RNA-IN in the antisense response.
Interestingly, of the predicted 15 unpaired residues in RNA-OUT, 8 exhibited increased reactivity to single strand-specific ribonucleases. While it is not obvious how to interpret this result, perhaps the simplest explanation is that Hfq binding to RNA-OUT prevents conversion of loops and/or bulges to structures that include base-paired regions.

Finally, a stretch of 5 consecutive residues in bulge 2 (nucleotides 23-27) exhibited a decrease in sensitivity to all 3 RNases (green asterisks). This suppression could result from Hfq binding to this segment. Notably, this is a very U-rich sequence (5’ UUAUUG 3’) that is predicted to be in single stranded form. *E. coli* Hfq has been shown to preferentially bind U-rich single stranded sequences in sRNAs through its proximal binding site (Ishikawa et al, 2012) and we provide evidence below that Hfq engages RNA-OUT exclusively through its proximal binding site (Figure 2.6). Using data from our RNase probing of RNA-OUT in the presence of Hfq, we determined a single structure of an RNA-OUT:Hfq complex (Figure S2.3A). This predicted structure is largely single stranded, with the exception of five base-pairs forming between nucleotides 29-33 and 59-63. Of the 35 nucleotides that are predicted to base-pair with RNA-IN, 30 are in single stranded regions after Hfq addition.

We also performed hydroxyl radical footprinting on Hfq:RNA-OUT complexes to further investigate the position(s) of Hfq binding (Figure 2.3A, lanes 4-7) but were unable to see clear and reproducible patterns of protection. The hydroxyl radical cleavage pattern was, however, useful in assigning cleavage products produced in the RNase structure-probe experiments.

We next probed the structure of the first 160 nucleotides of RNA-IN with RNases as described for RNA-OUT. In the absence of Hfq there was a significant amount of reactivity to V1 nuclease (compare lanes 2 and 11 in Figure 2.4A), which is indicative of this RNA forming base-paired segments. Two regions in which V1-reactive residues clustered (residues 17-35 and 45-60; indicated by a solid blue line in Figure 2.4A) also showed minimal reactivity to single strand-specific ribonucleases. Hard constraints from the nuclease data (circled letters in Figure 2.4B) were input into the Mfold program to generate a model of RNA-IN-160 structure. Notably, addition of these hard constraints
resulted in the output of a single structure. The model predicts one substantial stem that includes 11 base-pairs and 2 bulges. The stem includes residues 25-36 on one strand and residues 45-60 on the other strand. At least one nucleotide in each of the 11 predicted base-pairs showed reactivity to V1. We note that our model for RNA-IN-160 includes some secondary structure involving the first 20 nucleotides. However the single base-pair between position 6 and 149 is unlikely to maintain the most 5’ and 3’ portions of the RNA in a stable secondary structure. The absence of stable secondary structure within the first 8 nucleotides of RNA-IN suggests that there is no structural impediment to initiating pairing with RNA-OUT. Of the first 35 nucleotides of RNA-IN-160 that are complementary to RNA-OUT, residues 25-35 are sequestered in a stem that may interfere with the antisense response.
Figure 2.4. RNase footprinting of RNA-IN.

(A) $^{32}$P-labeled RNA-IN-160 (45 nM) was incubated with or without Hfq as indicated before treatment with ribonuclease A, T1, or V1 (lanes 5-13). Reactions, including RNA not treated with RNase (lanes 2-4) and a G-ladder (lane 1), were analyzed on a 10% denaturing polyacrylamide gel. Nucleotide labeling is relative to the RNA-IN in vitro transcriptional start site, which is nucleotide 1. Blue bars highlight clusters of V1 sensitivity observed in the absence of Hfq. (B) A model is shown for the secondary
structure of RNA-IN-160. The model was produced using Mfold with hard constrains (circled positions) obtained from two independent structure-probe experiments (part A and Figure S2.2). RNase A/T1 cleavage is indicated with red letters, while V1 cleavage is indicated with blue letters. Symbols (triangles and asterisks) are defined in the text.
Addition of Hfq to RNA-IN-160 had substantial effects on RNA structure. Much of the V1 sensitivity in the predicted stem was lost (indicated by downward pointing blue triangles in both Figures 2.4A and B). Strikingly, starting at position 104 and continuing to position 149 there was a large increase in V1 sensitivity with the addition of Hfq. As there were few residues within the 104-149 segment that showed substantial increases in either A or T1 sensitivity upon Hfq addition, it appears likely that some intramolecular base-pairing is occurring in this region. There were also a few regions that showed a decrease in reactivity to both single and double strand-specific ribonucleases. This includes segments 3-13, 17-25, 38-40 and 71-92 (denoted by green asterisks). These regions could define Hfq binding sites. Notably, at the highest concentration of Hfq used in this experiment, we anticipate, based on EMSA data (Figure S2.1B), that there could be as many as three distinct Hfq binding sites in RNA-IN-160. A predicted structure of RNA-IN-160 in the presence of Hfq is presented in Figure S2.3B. Consistent with the RNase footprinting data, the first 98 nucleotides of RNA-IN are mostly single stranded, with two small hairpins formed at nucleotides 34-56 and 66-79. An extensive stem-loop structure is predicted to form from nucleotides 99-160. We note that 33 out of 35 nucleotides of RNA-IN that are expected to base-pair with RNA-OUT are single-stranded in the presence of Hfq.

Note that RNase data for RNA-IN-160 footprinting comes from two independent experiments (Figure 2.4A and Figure S2.2). All reactivities were reproducible except at positions C72-C75 where we saw Hfq-dependent protection of these residues only in the experiment shown in Figure S2.2.

To further probe the location of Hfq binding sites in RNA-IN-160 we performed hydroxyl radical footprinting on 5’ end-labeled RNA-IN-160 using multiple Hfq concentrations (Figure 2.5). Quantitation of portions of the gel image showing the greatest differences in band intensity for ‘no Hfq’ (lane 2) and selected ‘plus Hfq’ samples (lanes 3, 6 and 10) is also presented. A region spanning residues 29 to 46 showed protection against hydroxyl radical cleavage at both 149 nM and 347 nM Hfq. At the higher Hfq concentration additional zones of protection were observed as indicated beside the gel image (green asterisks). Due to discontinuities in the patterns of protection
it is difficult to infer the boundaries of individual binding sites and therefore the total number of sites. However, based on the density of protected residues in the 29-46 segment and the fact that there was uniform protection in this cluster at an intermediate Hfq concentration, we suggest that this cluster defines a single Hfq binding site. The region spanning residues 84 to 94 includes the second highest density of protected residues and could represent a second Hfq binding site. This second site would be a lower affinity site relative to the site within the nucleotide 29-46 segment, as protections in this site were only observed at the highest Hfq concentrations. Unfortunately we were not able to obtain high quality hydroxyl radical footprinting data for the most 5’ portion of RNA-IN-160 to further test the possibility raised by RNase footprinting that this segment also contains an Hfq binding site. We do note that 5 residues (38, 39, 40, 45 and 46) within segment 29-46 showed a general suppression of cleavage by RNases (Figure 2.4A), as did residues 80, 82 and 86, which are close to or within segment 84-94, supporting the possibility that these segments include Hfq binding sites.
Figure 2.5. Hydroxyl radical footprinting of RNA-IN.

(A) $^{32}$P-labeled RNA-IN-160 (45 nM) was incubated with increasing concentrations of Hfq (lanes 3-10) and then subject to hydroxyl radical treatment (lanes 2-10). Lane 1 contains RNA not treated with hydroxyl radicals. Samples were analyzed as in Figure 2.4. Nucleotides are numbered as in Figure 2.4. Green asterisks identify positions protected from hydroxyl radical cleavage in the presence of Hfq, while purple asterisks identify positions where Hfq induced hypersensitivity to hydroxyl radical cleavage. (B) Quantification of band intensities from selected lanes of the gel image in part A is shown. Reactivity is presented in arbitrary units.
2.2.3 RNA-binding sites in Hfq that interact with RNA-IN and RNA-OUT

Hfq has at least two distinct RNA-binding surfaces, enabling it to simultaneously bind multiple RNAs and catalyze trans-sRNA/mRNA pairing reactions. To gain insight into how Hfq interacts with RNA-IN and RNA-OUT, we performed experiments designed to define the surfaces in Hfq that interact with these RNAs. It should be recognized that Hfq binding determinants for an asRNA have not previously been reported. Two complementary approaches were used. In one approach we performed binding assays with RNA-IN-160 or RNA-OUT and Hfq mutants that are defective in either proximal site (Hfq\textsubscript{K56A}) or distal site (Hfq\textsubscript{Y25A}) RNA binding. In the second approach we performed binding assays with Hfq\textsubscript{WT} and RNA-IN-160 or RNA-OUT in the presence of competitor RNAs that exhibit high affinities for either the proximal (DsrA) or distal (A\textsubscript{18}) RNA-binding surfaces of Hfq.

For binding experiments with Y25A and K56A mutant forms of Hfq, it was necessary to use Hfq bearing a C-terminal his\textsubscript{6} epitope tag (hereafter referred to as ‘his\textsubscript{6}-Hfq’). This is because, unlike Hfq\textsubscript{WT}, the two mutant forms of Hfq are not heat stable and therefore cannot be purified in the same way as untagged Hfq\textsubscript{WT} (Mikulecky et al, 2004). However, all three forms of the his\textsubscript{6}-tagged Hfq can be purified using nickel affinity chromatography. We show in Figure 2.6A and Table 2.1 that WT and Y25A forms of his\textsubscript{6}-Hfq bound RNA-OUT with similar affinities (K\textsubscript{D1} \approx 76 nM and 94 nM, respectively). In contrast, his\textsubscript{6}-Hfq\textsubscript{K56A} bound very poorly to RNA-OUT at Hfq concentrations up to 309 nM. At the high end of the Hfq\textsubscript{K56A} titration (upwards of 464 nM Hfq\textsubscript{K56A}) essentially all of RNA-OUT was bound by Hfq\textsubscript{K56A}, forming Hfq:OUT-2 with an apparent K\textsubscript{D} \approx 389 nM. This represents a 5-fold reduction in the affinity of his\textsubscript{6}-Hfq\textsubscript{K56A} versus his\textsubscript{6}-Hfq\textsubscript{WT} for RNA-OUT.
Figure 2.6. RNA-OUT interacts specifically with the proximal RNA-binding surface of Hfq.

(A) EMSAs with $^{32}$P-labeled RNA-OUT (~0.4 nM) and either WT or mutant forms of Hfq. Hfq$^{Y_{25}A}$ is defective in RNA-binding at the distal site, and Hfq$^{K_{56}A}$ is defective in RNA-binding at the proximal site. The corresponding curves are presented below each gel image. Error bars represent standard error from two experiments. Note that all forms of Hfq used in this experiment possess a his$_6$ epitope tag at their C termini. Species are labeled as in Figure 2.2. (B) EMSAs performed in the presence of competitor RNAs. Hfq$^{WT}$ (untagged) was first mixed with various concentrations of DsrA or A$^{18}$ RNA for 5 min, and $^{32}$P-labeled RNA-OUT (0.4 nM) was added. After an additional 15 min, reactions were subjected to polyacrylamide gel electrophoresis. A species expected to
represent a ternary complex is labeled $A_{18}:Hfq:OUT^*$. $IC_{50}$ values were calculated from curves shown in Figure S2.4 and are reported in Table 2.2.
The above results are consistent with Hfq binding RNA-OUT through its proximal site, which is typical of how Hfq binds trans-sRNAs. If this is correct, then it is expected that an sRNA, but not a distal-specific RNA, would act as a competitor for RNA-OUT binding to untagged Hfq\textsubscript{WT}. We show in Figure 2.6B that this is the case. When we pre-incubated Hfq with DsrA (an sRNA) or A\textsubscript{18} (a distal-specific RNA) and then added \textsuperscript{32}P-labeled RNA-OUT, only DsrA inhibited Hfq:OUT-1 complex formation; IC\textsubscript{50} values (Table 2.2 and Figure S2.4) are \( \sim 7 \) nM and \( > 4000 \) nM for DsrA and A\textsubscript{18}, respectively. In fact, there was evidence of ternary complex formation at A\textsubscript{18} concentrations above 31 nM (see lanes 18-24 in Figure 2.6B). Overall, we conclude that RNA-OUT behaves like a \textit{trans}-sRNA in its interaction with Hfq.

In titrations with RNA-IN-160 and his\textsubscript{6}-tagged Hfq proteins (WT, K56A and Y25A) there was a moderate reduction in binding affinity. This is reflected by changes in \( K_D \) of 1.7 and 3-fold, respectively for Hfq\textsubscript{Y25A} and Hfq\textsubscript{K56A} versus Hfq\textsubscript{WT} (see Figure 2.7A and Table 2.1). These results are consistent with RNA-IN-160 binding to both the proximal and distal sites. Competition experiments support this inference as both DsrA and A\textsubscript{18} gave some inhibition of Hfq:IN-1 formation when each of these RNAs was pre-incubated with Hfq prior to addition of RNA-IN-160. More specifically, we observed strong inhibition of Hfq:IN-1 formation at concentrations above the \( K_D \) for A\textsubscript{18}:Hfq complex formation, which is \( \sim 10 \) nM (Sun & Wartell, 2006) (Figure 2.7B, lanes 11-15; IC\textsubscript{50} \( \sim 11 \) nM, Table 2.2; Figure S2.4). In contrast, we observed weak inhibition of Hfq:IN-1 formation at DsrA concentrations above the \( K_D \) for DsrA:Hfq complex formation, which is \( \sim 21 \) nM (Mikulecky et al, 2004) (Figure 2.7B, lanes 6-8; IC\textsubscript{50} \( \sim 53 \) nM – Table 2.2 and Figure S2.4). At A\textsubscript{18} concentrations above its \( K_D \) for Hfq (lanes 11-15) primarily the proximal site is expected to be available for RNA-IN-160 binding and at DsrA concentrations above its \( K_D \) for Hfq (lanes 5-7) primarily the distal site is expected to be available for RNA-IN-160 binding. Accordingly, the stronger inhibition observed for A\textsubscript{18} is consistent with the distal site of Hfq being the higher affinity site for RNA-IN-160 binding. This is typical of Hfq binding to mRNAs (Mikulecky et al, 2004; Soper & Woodson, 2008).
We also performed a competition experiment where both competitors were mixed with Hfq simultaneously and then RNA-IN-160 was added (Figure 2.7B, lanes 16-26). Very strong inhibition of Hfq:IN-1 formation was only observed when the concentrations of both A_{18} and DsrA were close to or above their respective K_D values for Hfq complex formation (lanes 23-26). These results suggest that an additional RNA binding site in Hfq does not contribute significantly to the formation of a stable Hfq:RNA-IN-160 complex.
Table 2.2. *In vitro* competition by DsrA or A$_{18}$ for binding of Hfq$^{WT}$ to RNA-IN or RNA-OUT

<table>
<thead>
<tr>
<th>RNA Species</th>
<th>DsrA IC$_{50}$ (nM)</th>
<th>A$<em>{18}$ IC$</em>{50}$ (nM)</th>
<th>DsrA+A$<em>{18}$ IC$</em>{50}$ (nM)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA-OUT</td>
<td>6.77 ± 0.317</td>
<td>&gt;4000$^b$</td>
<td>N.D.</td>
</tr>
<tr>
<td>RNA-IN</td>
<td>52.7 ± 14.1</td>
<td>10.6 ± 1.18</td>
<td>8.93 ± 2.12</td>
</tr>
</tbody>
</table>

IC$_{50}$ values (± standard error) were measured by EMSA. Percentage of competition was plotted vs competitor concentration, and IC$_{50}$ values were obtained from the resulting curves (shown in Figure S2.4). Competition assays were performed.

N.D., not determined.

$^a$Reported for each competitor in the mix.

$^b$Instead of competition, a ternary complex was formed.
Figure 2.7. RNA-IN interacts with the distal and proximal RNA-binding surfaces of Hfq.

(A) EMSAs with $^{32}$P-labeled RNA-IN (0.17 nM) and either WT or mutant forms of his$_6$-tagged Hfq. Species are labeled as in Figure 2.2. Binding curves are shown below the corresponding EMSA, and the apparent $K_D$ values are reported in Table 2.1. Error bars represent standard error from two experiments. (B) EMSAs performed in the presence of competitor RNAs. Competitor experiments were performed as described in Figure 2.6B except RNA-IN was present at a concentration of 0.17 nM. For lanes 18-26, a 1:1 mix of DsrA and A$_{18}$ was serially diluted to the indicated concentrations before competition. IC$_{50}$ values were calculated from curves shown in Figure S2.4 and are reported in Table 2.2.
2.2.4 Hfq\textsubscript{K56A} exhibits a reduced rate of RNA-IN:RNA-OUT pairing \textit{in vitro}

The results in the previous section show that RNA-IN can contact both the distal and proximal sites on Hfq and that RNA-OUT binds only to the proximal site. One or more of these interactions is likely important for the acceleration in the rate of RNA-IN:OUT pairing directed by Hfq that we previously documented (Ross et al, 2010). The proximal site is likely of particular importance, as both RNA species could conceptually bind here and begin to pair. Accordingly, we asked if an intact proximal RNA-binding site is necessary for Hfq to accelerate the rate of RNA-IN:OUT pairing. Note that we developed an alternative means to purify untagged Hfq\textsubscript{WT} and Hfq\textsubscript{K56A} (see Materials and Methods), as the his\textsubscript{6}-tagged forms gave inconsistent pairing results. We mixed Hfq (WT or K56A), \textsuperscript{32}P-labeled RNA-OUT and \textsuperscript{32}P-labeled RNA-IN-160 and incubated them for the indicated times before processing and analysis on a native polyacrylamide gel. Processing involved phenol extraction (mixing, centrifugation and recovery took ~15 seconds) and then immediately loading the samples on a native polyacrylamide gel. This step was included to remove Hfq from the paired product so that identification of this species was unambiguous. The ratio of RNA-OUT to RNA-IN was fixed (10:1) to roughly reflect the ratio of these RNAs \textit{in vivo}. The Hfq concentration used (45 nM) was set from a preliminary experiment where we determined the minimum concentration of Hfq that yielded an enhancement in RNA-IN:OUT pairing (relative to the absence of Hfq) at the above ratio of RNA-OUT to RNA-IN (see Figure S2.5).

We show in Figure 2.8 that addition of 45 nM Hfq\textsubscript{WT} increased the rate ($k\textsubscript{obs}$) of RNA-IN:OUT pairing approximately 19-fold relative to no Hfq addition ($k\textsubscript{obs} = 1.51 \text{ min}^{-1}$ for Hfq\textsubscript{WT} \textit{versus} 0.080 min\textsuperscript{-1} in the absence of Hfq). By comparison, the rate enhancement was less than 2-fold when Hfq\textsubscript{K56A} was used ($k\textsubscript{obs} = 0.14 \text{ min}^{-1}$). These results indicate that the K56A mutation negatively impacts RNA-IN:OUT pairing under these specific conditions, consistent with the proximal surface playing an important role in the enhancement of antisense pairing \textit{in vitro}. 
Figure 2.8. RNA-IN:RNA-OUT pairing reactions.

(A) $^{32}$P-labeled RNA-IN (0.85 nM) was mixed with excess $^{32}$P-labeled RNA-OUT (8.5 nM) and, where indicated, untagged WT or K56A Hfq (45 nM). At the indicated time points, pairing reactions were stopped by treatment with a phenol/water mix and immediately loaded onto a 6% native polyacrylamide gel. (B) The amount of RNA-OUT:RNA-IN complex (OUT:IN) was determined as a percentage of total RNA-IN for each time point and plotted as a function of time. Error bars represent the standard error from three experiments. The observed rate constant ($k_{obs}$) is indicated for each reaction. These values were derived from curves corresponding to the equation describing the rate of exponential association, presented in Materials and Methods.
2.2.5 IS10-kan transposition is derepressed in strains expressing Y25A and K56A forms of Hfq

While Hfq increases the rate of RNA-IN:OUT pairing substantially \textit{in vitro}, it is difficult to know if effective antisense inhibition of transposase expression requires this enhancement in the RNA pairing rate \textit{in vivo}. To address this issue we asked if untagged Hfq\textsubscript{K56A} (which is impaired in its ability to promote IN:OUT pairing \textit{in vitro}) is also less effective than wild type Hfq at repressing IS10 transposition. We also assessed the ability of the distal-impaired Hfq\textsubscript{Y25A} to repress transposition of IS10. We used a ‘mating out’ assay to measure the frequency of IS10-Kan transposition from a multi-copy plasmid in different genetic backgrounds. In this assay, the frequency of transposition of IS10-Kan from a multi-copy plasmid to the F plasmid in the donor strain is measured (see Materials and Methods). For the mating out experiments, the Hfq status was manipulated by transforming the donor strain (DBH16; hfq\textsuperscript{−}) with a plasmid expressing untagged WT, Y25A or K56A forms of Hfq from a native hfq promoter. As controls, we also measured the frequency of IS10-Kan transposition in hfq\textsuperscript{+} (DBH33; full repression) and hfq\textsuperscript{−} (no repression) strains.

The results of the mating out analysis are presented in Figure 2.9 where we report the fold-change in transposition frequency relative to the average transposition frequency calculated for hfq\textsuperscript{−}. In both the hfq\textsubscript{K56A} and hfq\textsubscript{Y25A} strains, IS10-Kan transposition was derepressed to about the same level as in hfq\textsuperscript{−}. Importantly, these deficiencies cannot be attributed to differential levels of Hfq expression in the different strains as Western Blotting confirmed that plasmid-encoded WT, Y25A and K56A forms of Hfq were present at comparable levels in the respective donor strains (Figure S2.6). These results show that Hfq mutants that are partially defective in binding RNA at specific surfaces, one of which (K56A) is impaired in its ability to catalyze RNA-IN:OUT pairing \textit{in vitro}, are unable to repress IS10 transposition \textit{in vivo}.

Hfq is also known to stabilize \textit{trans}-encoded sRNAs (reviewed in (Vogel & Luisi, 2011)) and such an activity could influence the effectiveness of RNA-OUT in the Tn10/IS10 antisense system. We performed a rifampicin time-course experiment to look at this possibility. We show in Figure S2.7 that the half-life of RNA-OUT actually
Figure 2.9. IS10-kan transposition is derepressed in *E. coli* encoding Y25A and K56A forms of Hfq.

*E. coli* cells (*hfq*<sup>+</sup> or *hfq*) were cotransformed with pDH602 (encodes IS10-kan) and a compatible plasmid encoding untagged Hfq (WT, K56A, or Y25A) or the corresponding “empty vector” control. Relative transposition frequencies were measured using the conjugal mating out assay. An average transposition frequency (4.03 x 10<sup>-3</sup> events per mL of mating mixture) was calculated for the *hfq*<sup>-</sup> strain (*hfq*<sup>-</sup>/emp.vect.) from 15 independent “donour” colonies across four independent experiments, and this was value was set to 1. All other transposition values are expressed relative to this value where Hfq-directed repression of transposition is at its maximal level. Bars indicate the mean; the error bars indicate standard error on the mean. From left to right, the n value for each treatment group is 15, 15, 14, 16, 11 – these were compiled from at least two (and up to four) independent experiments. An asterisk (*) indicates that means were significantly different from the *hfq*<sup>-</sup> control group; p-values are indicated above the corresponding bars.
decreased by about 2.5-fold in an \textit{hfq}+ relative to an \textit{hfq}- strain. Accordingly, we can rule out the possibility that Hfq contributes to the antisense system by stabilizing RNA-OUT.

2.3 Discussion

Tn10/IS10 transposition is negatively regulated by an asRNA (RNA-OUT) that pairs with the 5' TIR of the transposase mRNA (RNA-IN) to inhibit transposase expression. Hfq is also a negative regulator of Tn10/IS10 transposition that down-regulates transposase expression \textit{in vivo}. The effect of disrupting \textit{hfq} on Tn10/IS10 transposition is diminished in a system with reduced RNA-OUT expression. \textit{In vitro}, Hfq binds RNA-IN, RNA-OUT, forms a ternary complex with these RNAs and increases the rate of RNA-IN:OUT pairing. Taken together, these observations are consistent with Hfq regulating Tn10/IS10 by operating on the antisense mechanism. In the current work we have further characterized the interactions between Hfq, RNA-IN and RNA-OUT with regard to binding affinity, binding sites within Hfq that govern these RNA contacts and the impact mutating one of these binding sites has on RNA-IN:OUT pairing. We have also begun to define Hfq-binding sites within each RNA, as well as the impact Hfq binding has on the structure of these RNAs.

2.3.1 RNA-IN and RNA-OUT bind Hfq like a prototypical sRNA-mRNA pair

The Hfq binding affinities we have measured for RNA-IN and RNA-OUT are consistent with what is typically seen for canonical trans-encoded sRNA-mRNA pairs. For example, $K_D$ for RNA-OUT is 19.6 nM and $K_D$ for Hfq binding to DsrA is 21 nM (Mikulecky et al, 2004). Hfq binds RNA-IN with sub-nanomolar affinity (apparent $K_D$ 0.24 nM), which is comparable to the tightest Hfq-mRNA interactions described to date (OmpC 0.9 nM, (Fender et al, 2010); SodB 0.3 nM, (Geissmann & Touati, 2004); RpsO 90 pM, (Folichon et al, 2003)).

We also investigated determinants in Hfq responsible for RNA-IN and RNA-OUT binding. Hfq possesses at least two distinct RNA-binding surfaces; the distal site
generally binds A-rich RNA/mRNA while the proximal site binds U-rich sRNAs (Ishikawa et al, 2012; Mikulecky et al, 2004; Soper et al, 2011). Both competition and binding experiments (with Hfq variants) revealed that Hfq binds RNA-OUT exclusively through its proximal RNA-binding surface. For RNA-IN, both the distal and proximal binding sites in Hfq contribute to RNA-IN binding, although the distal site is the higher affinity site. In canonical sRNA-mRNA systems there is competition between sRNA and mRNA binding at the proximal site and this appears to be required for the formation of the paired species (Hwang et al, 2011). Competition is ensured by the individual RNAs having similar binding affinities for the proximal site and in general these affinities are much weaker than those for the distal site. The high affinity interaction for the mRNA with the distal site effectively tethers the mRNA to Hfq, allowing other parts of the mRNA to interact relatively weakly with the proximal site and this increases the probability that the mRNA and sRNA can occupy the proximal site at the same time (Hopkins et al, 2011). As discussed above the Hfq:RNA interactions in the IS10 system are consistent with this general model. A recent study defined a third RNA-binding site in Hfq (the lateral surface) that may be important in allowing the mRNA to simultaneously bind distal and proximal sites (Sauer et al, 2012). Our competition studies indicate that the lateral surface alone is not sufficient for binding RNA-IN or RNA-OUT.

We have not defined the number of Hfq hexamers present in any of the Hfq:RNA-IN or Hfq:RNA-OUT complexes. The $K_D$ values discussed above relate only to the complexes formed at the lowest Hfq concentrations in each titration. For RNA-IN, at least four distinct Hfq complexes were detected, raising the possibility that at least four Hfq hexamers may be accommodated within the first 160 nucleotides of RNA-IN. Results from hydroxyl radical footprinting support the existence of multiple Hfq binding sites in RNA-IN. One such site may extend from position 29-46 and appears to be the highest affinity site identified by hydroxyl radical footprinting, as it was occupied at a lower Hfq concentration relative to the other sites. For RNA-OUT, two distinct Hfq-bound species were detected. Results from RNase footprinting revealed only one strong candidate for an Hfq binding site within RNA-OUT. This site is located within the U-rich segment of bulge 2. We do not yet know which Hfq binding site(s) in RNA-IN or OUT are biologically relevant. Work in other systems is consistent with the idea that maximal
pairing of an mRNA:sRNA pair can require the mRNA to bind multiple Hfq hexamers (Salim & Feig, 2010; Soper & Woodson, 2008). Mutagenesis of potential Hfq-binding sites in RNA-IN and RNA-OUT is currently under way to test the importance of individual sites in IS10 transposition.

2.3.2 A proximal site mutation impacts on RNA-IN:OUT pairing and IS10 transposition

The rate of RNA-IN:OUT pairing was substantially enhanced by HfqWT but not HfqK56A. A limitation of this analysis was that we measured \( k_{\text{obs}} \) under one specific set of parameters as opposed to measuring a second order rate constant. We used excess RNA-OUT relative to RNA-IN (10:1), and a small excess of Hfq relative to both (less than 5-fold). As ‘available’ Hfq is thought to be limiting in the cell (Hussein & Lim, 2011; Moon & Gottesman, 2011) we feel this is a reasonable approximation of \textit{in vivo} conditions. Under these conditions, the \( k_{\text{obs}} \) value was 10.6-fold lower for HfqK56A versus HfqWT.

We also measured the impact of the Hfq proximal and distal site mutations on IS10 transposition. IS10 transposition was derepressed in both \( hfq_{K56A} \) and \( hfq_{Y25A} \) strains relative to \( hfq_{WT} \). RNA-IN:OUT pairing \textit{in vitro} was enhanced \( \sim19 \)-fold by HfqWT relative to no Hfq, and the magnitude of de-repression of IS10 transposition in \( hfq^- \) relative to \( hfq^+ \) was \( \sim10\text{-}15\)-fold. Furthermore, \textit{in vitro} pairing was \( \sim10\)-fold faster in the presence of HfqWT relative to HfqK56A, and \textit{in vivo} transposition was de-repressed \( \sim12\)-fold for \( hfq_{K56A} \) relative to plasmid-borne \( hfq^+ \). Taken together these results are consistent with Hfq playing a significant role in the pairing component of the IS10 antisense system. Another way in which Hfq might facilitate the IS10 antisense system is through the stabilization of RNA-OUT. However, this possibility is not supported by our observation that RNA-OUT stability is actually reduced in an \( hfq^+ \) compared to an \( hfq^- \) strain. Finally, Hfq might directly interfere with IS10 transposase translation. We think this is unlikely because as previously noted the large increase in Tn10/IS10 transposition from a multi-copy plasmid in \( hfq^- \) is tightly linked to the expression of RNA-OUT.
2.3.3 How might Hfq promote pairing in the IS10 antisense system?

The simplest scenario for how Hfq promotes RNA pairing in the IS10 antisense system is that, through simultaneous binding of RNA-IN and RNA-OUT, Hfq acts as a pairing catalyst by increasing the local concentration of these two RNAs. In addition to providing a single surface to which both RNAs bind, Hfq might actively alter the structure of RNA-IN and RNA-OUT to promote pairing. In fact, our structure-probing data supports an Hfq-dependent restructuring model (Figure 2.10). On the left-hand side of Figure 2.10 we show how RNA-IN and RNA-OUT might interact in the absence of Hfq. Structure probing experiments with RNA-IN and RNA-OUT revealed that the pairing region of both RNAs is at least partly sequestered in secondary structure. In this pathway a total of 8 base-pairs between RNA-IN and RNA-OUT could readily form without any RNA restructuring (structure vii). In the right-hand panel we show how pairing could occur with RNAs (structures iii and vi) that have been restructured by Hfq. In the presence of Hfq the pairing region of both RNAs is largely single-stranded leading to the formation of a paired species (structure viii) that contains 30 base-pairs between RNA-IN and OUT. Notably, only in structure viii is the TIR sequestered through base-pairing with RNA-OUT. We anticipate that upon Hfq dissociation, RNA-OUT will adopt its native structure (transition from structure iii to structure i). However, the capacity of the 3’ end of RNA-IN to form a stable secondary structure in the presence of Hfq may prevent Hfq-bound RNA-IN (structure vi) from converting back to structure iv upon Hfq release. As Hfq rapidly cycles on and off of RNAs (Fender et al, 2010) the formation of the 3’ stem-loop structure could maintain RNA-IN in a ‘pairing competent’ state after Hfq release. Consistent with genetic data, the ‘Hfq pairing pathway’ includes pairing of the 5’ terminus of RNA-IN with the hairpin loop of RNA-OUT. Also, the internal loop of RNA-OUT has been shown to be important for the antisense response (Case et al, 1989; Jain, 1995; Kittle et al, 1989). In the Hfq-independent pathway nucleotides within this loop can directly pair with RNA-IN facilitating further propagation of pairing. In the Hfq pathway we suggest that internal loop residues directly participate in Hfq binding and are therefore important in the restructuring of RNA-OUT.
Figure 2.10. Model for RNA-IN:OUT antisense pairing in the presence vs. absence of Hfq.

The Hfq-independent pairing pathway is shown on the left-hand side (structures i, iv, and vii) and the Hfq-dependent pathway is shown on the right-hand side (structures iii, vi, viii). In structures (ii) and (v) Hfq is shown bound to RNA-OUT and RNA-IN, respectively, but conformational changes in the RNAs have not yet taken place. Other structures are described in the text. Hfq hexamers are indicated by green circles. The start codon (AUG) and Shine-Dalgarno sequence (SD) are indicated by asterisks (*) at the first
nucleotide of each sequence. Intramolecular base-pairs in RNA-OUT/IN are indicated by blue and red, respectively. Intermolecular base-pairs between RNA-OUT and RNA-IN are in grey.
An active remodeling role for Hfq has been well documented in other systems. Hfq alters the structure of RpoS mRNA such that the sequence that base-pairs with DsrA becomes single stranded (Soper et al, 2011). The pairing region within SodB mRNA is also sequestered in a stem-loop structure; Hfq binding is required to disrupt this stem-loop, resulting in formation of a large loop that is then competent to pair with the sRNA, RyhB (Geissmann & Touati, 2004).

In other systems the presence of base-pairing discontinuities in structured regions of asRNAs has also been shown to be critical for antisense regulation (see (Wagner et al, 2002)). At this point it is unclear if these discontinuities are sufficient for a robust antisense response in vivo and in this regard it will be interesting to see if the effectiveness of other antisense systems shows any reliance on Hfq.

In the current work we have provided additional insight into how Hfq interacts with the RNA components of the IS10 antisense system and provided further evidence that these interactions ultimately influence this system. To date only one other antisense system is known to be Hfq-regulated. Hfq regulates the expression of the chromosomally encoded gadX gene, which is involved in acid tolerance in E. coli. An sRNA called GadY is antisense to the 3’ UTR of GadX and it has been shown that GadY expression increases the stability of the GadX transcript. Importantly, Hfq binds to the GadY transcript and stabilizes it (Opdyke et al, 2004). However, it has not been established if Hfq plays a direct role in promoting pairing of GadY and GadX RNAs. With regard to other transposons it should be noted that antisense RNAs to the transposase have been identified, including IS30 (Arini et al, 1997) and IS200 (Sittka et al, 2008). Additionally, a recent study identified 5 transcripts in Mycobacterium smegmatus that bind Hfq and are antisense to transposase mRNAs (Li et al, 2013). It will be interesting to see if any of these transposons are regulated by Hfq. Notably, Hfq regulation of other transposons might not be limited to systems encoding asRNAs. Trans-encoded sRNAs frequently target more than just one mRNA (reviewed in (Repoila et al, 2003)) and as such there is the potential for ‘off-target’ effects wherein a trans-encoded sRNA might fortuitously target a transposase mRNA. This could provide the host with a previously unrecognized pathway to either down- or up-regulate transposon mobilization. Furthermore, as the
induction of the transcription of sRNA genes if often linked to various cellular stresses (Repoila et al, 2003) and this induction can temporarily limit the availability of Hfq in the cell (Hussein & Lim, 2011; Moon & Gottesman, 2011), there is the potential to indirectly regulate transposition reactions through stress response pathways.

2.4 Materials and Methods

2.4.1 Strains, plasmids, and primers

All *E.coli* strains and plasmids used in this study are listed in Table 2.3. Oligonucleotides used are listed in Table 2.4.

To express Hfq *in vivo* we cloned a fragment that included the *hfq* gene with its *P3* promoter into the low-copy cloning vector pWKS30 (Wang & Kushner, 1991) The aforementioned fragment was generated by PCR using genomic DNA from DBH33 and primers JR15 and JR16, which include XbaI and HindIII sites, respectively. After digestion of the PCR product with XbaI-HindIII, the ‘*hfq*’ fragment was ligated into XbaI-HindIII digested pWKS30, creating pDH700. We then used pDH700 as a template for site-directed mutagenesis to create pDH701 and pDH713, which encode Hfq_{K56A} and Hfq_{Y25A}, respectively. For purification of C-terminal his_{6}-tagged Hfq (WT, Hfq_{Y25A} and Hfq_{K56A}) the *hfq* gene was cloned into pET28a as described in (Mikulecky et al, 2004).

2.4.2 Hfq purification and quantitation

Untagged Hfq was purified as described in (Ross et al, 2010), but included a treatment of the lysate with DNase I (100 units) and RNase A (100 µg) for 1 hour on ice before heat treatment of the lysate at 85°C. His_{6}-Hfq proteins were purified as described in (Mikulecky et al, 2004). Untagged Hfq_{WT} and Hfq_{K56A} for the experiments presented in Figure 2.8 were expressed from pDH700 and pDH701 and purified on a TALON Cobalt column as described in (Soper et al, 2010), followed by further purification on a polyA column as described in (Ross et al, 2010). Purified Hfq was dialyzed against Hfq storage/binding buffer (50 mM Tris, pH 7.5, 1 mM EDTA, 250 mM NH_{4}Cl, 10% (v/v)
Table 2.3. List of *E. coli* strains and plasmids used in Chapter 2.

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Relevant genotype</th>
<th>Use</th>
<th>Source/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HB101</td>
<td>F- leu' ; Str&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Mating out recipient</td>
<td>(Bolivar &amp; Backman, 1979)</td>
</tr>
<tr>
<td>DBH33</td>
<td>NK5830; recA&lt;sup&gt;arg&lt;/sup&gt; / F'&lt;sup&gt;pro&lt;/sup&gt;</td>
<td>Mating out donor</td>
<td>(Ross et al, 2010)</td>
</tr>
<tr>
<td>DBH16</td>
<td>NK5830 hfq-1::Ω&lt;sup&gt;cat&lt;/sup&gt;; Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Mating out donor</td>
<td>(Ross et al, 2010)</td>
</tr>
<tr>
<td>DBH5α BL21</td>
<td>recA&lt;sup&gt;-&lt;/sup&gt; / DE3 T7 RNA polymerase</td>
<td>Plasmid propagation</td>
<td>Invitrogen (Studier &amp; Moffatt, 1986)</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pDH602</td>
<td>pACYC184-derived; IS10-Kan ; Cm&lt;sup&gt;R&lt;/sup&gt;Kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Mating out assays</td>
<td>(Ross et al, 2010)</td>
</tr>
<tr>
<td>pDH631</td>
<td>pET3a derived; T7-hfq ; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Hfq over-expression</td>
<td>(Ross et al, 2010)</td>
</tr>
<tr>
<td>pWK30</td>
<td>pSC101-derived; low copy-number ori ; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Empty vector for Hfq expression</td>
<td>(Wang &amp; Kushner, 1991)</td>
</tr>
<tr>
<td>pDH700</td>
<td>pWKS30-P3-hfq&lt;sub&gt;WT&lt;/sub&gt; ; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Hfq&lt;sub&gt;WT&lt;/sub&gt; expression</td>
<td>This study</td>
</tr>
<tr>
<td>pDH701</td>
<td>pWKS30-P3-hfq&lt;sub&gt;K56A&lt;/sub&gt; ; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Hfq&lt;sub&gt;K56A&lt;/sub&gt; expression</td>
<td>This study</td>
</tr>
<tr>
<td>pDH713</td>
<td>pWKS30-P3-hfq&lt;sub&gt;Y25A&lt;/sub&gt; ; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Hfq&lt;sub&gt;Y25A&lt;/sub&gt; expression</td>
<td>This study</td>
</tr>
<tr>
<td>pDH686</td>
<td>pET28a-derived; C-terminal his&lt;sub&gt;6&lt;/sub&gt;-tagged Hfq&lt;sub&gt;WT&lt;/sub&gt;; kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Hfq&lt;sub&gt;WT&lt;/sub&gt; overexpression</td>
<td>(Mikulecky et al, 2004)</td>
</tr>
<tr>
<td>pDH688</td>
<td>pET28a-derived; C-terminal his&lt;sub&gt;6&lt;/sub&gt;-tagged Hfq&lt;sub&gt;K56A&lt;/sub&gt;; kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Hfq&lt;sub&gt;K56A&lt;/sub&gt; overexpression</td>
<td>(Mikulecky et al, 2004)</td>
</tr>
<tr>
<td>pDH697</td>
<td>pET28a-derived; C-terminal his&lt;sub&gt;6&lt;/sub&gt;-tagged Hfq&lt;sub&gt;Y25A&lt;/sub&gt;; kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Hfq&lt;sub&gt;Y25A&lt;/sub&gt; overexpression</td>
<td>(Mikulecky et al, 2004)</td>
</tr>
</tbody>
</table>
### Table 2.4. List of oligonucleotides used in Chapter 2.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’ to 3’)</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>JR1</td>
<td>TAATACGACTCAAATAGGCGAAAAATCAATAATCGAGACAAACAAG</td>
<td>Forward primer (includes T7 promoter) for <em>in vitro</em> transcription template (RNA-IN)</td>
</tr>
<tr>
<td>JR2-2</td>
<td>CAAGTTCCGGTAAGAGTGAGAG</td>
<td>Reverse primer for above (RNA-IN-160)</td>
</tr>
<tr>
<td>JR3</td>
<td>GAAATGTAAAACTCAGACTATAGGCTGCGCTGGACTTAGGCTGCAACATCTTGTTGTC</td>
<td>Forward primer (includes T7 promoter) for <em>in vitro</em> transcription template (RNA-OUT)</td>
</tr>
<tr>
<td>JR4</td>
<td>GGATACACATCTTGTCACTATGATCA</td>
<td>Reverse primer for above</td>
</tr>
<tr>
<td>JR5</td>
<td>GACATGGGTGTATTCATCCGCGACGCGATTTC</td>
<td>Reverse primer for above</td>
</tr>
<tr>
<td>JR6</td>
<td>NNAAGCTTNNTATTGATTTCTTGCCTGTATATGCA</td>
<td>Reverse primer for above</td>
</tr>
<tr>
<td>JR7</td>
<td>AGCCAGATACGGGTACCGCGACGCGATTTC</td>
<td>Forward primer for K56A quikchange mutagenesis</td>
</tr>
<tr>
<td>JR8</td>
<td>AGTAGAAATCGGCTGCGCGTGAAACATCTGGCT</td>
<td>Reverse primer for above</td>
</tr>
<tr>
<td>JR9</td>
<td>GCTGATGGTTGCTTATGGAAGTGGTGAATGGTATAGGA</td>
<td>Reverse primer for above</td>
</tr>
<tr>
<td>JR10</td>
<td>CTTAATACCGCAGCCAAAGCAATAGGAACTGGCTGGACACGACATCA</td>
<td>Forward primer for <em>in vitro</em> transcription template (DsrA)</td>
</tr>
<tr>
<td>JR11</td>
<td>AAATCCCGACCCCGCTGGGAGGTCGGGAT</td>
<td>Reverse primer for above</td>
</tr>
</tbody>
</table>
glycerol). SDS-PAGE revealed that the Hfq was ~ 95% pure (Figure S2.6B). Hfq concentration was determined by Bradford assay.

2.4.3 In vitro transcription and RNA purification

Linear DNA templates for run-off transcription of RNA-IN (nucleotides 1-160) or RNA-OUT (nucleotides 1-69) were amplified from pDH602 by PCR with primers JR1/JR2-2 or JR3/JR4 respectively; note that for each primer pair the forward primer includes the T7 core promoter. The same approach was used to make templates for in vitro transcription of DsrA (primers JR21/JR22). Our standard in vitro transcription reaction for generating unlabeled RNA was performed in a 30 µL volume with 200 ng DNA template, 2.5 mM rNTPs, 10 mM DTT, 1X T7 RNA polymerase reaction buffer (NEB), 100 U RNasin (Promega), 2.5 U yeast inorganic pyrophosphatase (NEB) and 100 U T7 RNA polymerase (NEB). For preparing 32P-labeled RNA, in vitro transcription was performed in a 20 µL volume as above except that UTP was added to only 50 nM, and 2.5 µCi [α-32P]UTP was added. Reaction mixtures were incubated at 37°C for 1 hour before adding 0.1 U of Turbo DNase (Ambion) per µL of reaction and continuing incubation for 20 minutes. RNAs were purified using denaturing PAGE and after elution from gel slices were concentrated by ethanol precipitation and finally re-suspended in Hfq storage/binding buffer. RNA concentrations were determined using a NanoSpectrophotometer (IMPLEN). Purity of in vitro transcribed RNA was assessed by high resolution denaturing PAGE. A18 was purchased from Sigma-Aldrich and dissolved in Hfq storage/binding buffer.

2.4.4 Hfq-RNA binding assays

In our standard Hfq-RNA binding reaction we mixed 32P-labeled RNA (0.1-0.4 nM) with Hfq (0.05-1856 nM) in Hfq storage/binding buffer (total reaction volume 10 µL) for 15 minutes at 37°C. In the case of ‘competitor’ experiments, unlabeled competitor RNAs (0.5-4000 nM) were mixed with Hfq for 5 minutes as described above and then either 32P-labeled RNA-IN (0.17 nM) or RNA-OUT (0.4 nM) was added. Incubation was continued at 37°C for an additional 15 minutes. Prior to mixing RNA with Hfq, the various RNA species were incubated at 95°C for 2 minutes, placed on ice for 2 minutes,
and equilibrated to 37°C. At the reaction end points samples were mixed with 0.3 volumes of gel load dye (21 mM Tris-HCl pH 7.5, 10 mM DTT, 100 mM KCl, 30% (v/v) glycerol, 0.05% (w/v) bromophenol blue) and applied to a 6% native polyacrylamide gel. Electrophoresis was carried out at 14 V/cm for 70 minutes whereupon the gel was dried and exposed to a phosphorimager screen. Gel images were obtained using the STORM phosphorimager.

For measuring the equilibrium dissociation constant (K_D) in binding reactions, bands representing shifted and unshifted RNA species were quantified (ImageQuant software) and the percentage of counts for a given shifted species (relative to total counts for all bands in the lane) was plotted on the y-axis and Hfq concentration on the x-axis (Prizm software). The resulting curve was fit by non-linear regression to the equation:

$$P:L = \frac{P:L_{max} \times [P]^h}{K_D^h + [P]^h}$$

where P and L are Hfq and RNA, respectively, P:L is the percentage of RNA shifted by Hfq, [P] is the concentration of Hfq in nM, K_D is the equilibrium dissociation constant and h is the Hill Coefficient. To calculate K_D1, the percentage of all shifted species were summed to yield the appropriate curve (e.g. the curve marked ‘total’ in Figure 2.2B,D). To calculate K_D2 for RNA-OUT, the percentage of all species representing complex 2 were plotted (e.g. the curve marked ‘Hfq:OUT*2’ in Figure 2.2B). To calculate K_D2 for RNA-IN-160, the total percentage of all species other than free RNA-IN-160 or Hfq:IN*1 were plotted (i.e. the curve marked ‘Hfq:IN*2’ in Figure 2.2D). To calculate K_D3 for RNA-IN-160, the total percentage of all species other than free RNA-IN-160, Hfq:IN*1 or Hfq:IN*2 were plotted (i.e. the curve marked ‘Hfq:IN*3 in Figure 2.2D).

For calculating IC_{50} values (i.e. the concentration of competitor RNA that inhibited Hfq:RNA-IN or Hfq:RNA-OUT complex formation by 50%), bands representing Hfq:RNA* complexes (* denotes radiolabelled RNA-IN or RNA-OUT) and unshifted RNA* were quantified and the percentage of Hfq-shifted RNA* at 0 nM competitor was set at 0% competition. The percentage of Hfq-shifted RNA* remaining at increasing concentrations of competitor was subtracted from 100% to give the percent competition
(0% shifted complex = 100% competition). Percent competition was plotted on the y-axis and competitor concentration on the x-axis (Prizm). The resulting curve was fit by non-linear regression to the equation:

\[ PC = \frac{PC_{\text{max}} \times [C]}{IC_{50} + [C]} \]

where PC is Percent Competition, [C] is the concentration of competitor RNA in nM and IC_{50} is the concentration of competitor RNA giving 50% competition.

2.4.5 RNA structure-probing and footprinting

*In vitro* transcribed RNA-IN and RNA-OUT were gel purified, treated with Antarctic Phosphatase (NEB) and 5’ end-labeled with \([γ-^{32}]\text{P}\)ATP (Perkin Elmer) and OptiKinase (USB). 5’-labeled RNA was gel purified, ethanol precipitated and finally re-suspended in RNA Storage Buffer (20 mM MES pH 7.0, 150 mM NaCl, 1 mM EDTA). RNA concentrations were determined using a NanoSpectrophotometer (IMPLEN). RNA and Hfq were mixed in RNA Structure Buffer (20 mM MES pH 7.0, 150 mM NaCl, 1 mM EDTA, 10 mM MgCl_{2}) to a final volume of 9 µL. Binding reactions took place at 37°C for 15 minutes. For RNase footprinting, 1 µL of dilute RNase A, T1, or V1 (Ambion) was added to each binding reaction. For RNA-OUT 0.04 ng RNase A, 0.03U or 0.04 U of RNase T1 (- or + Hfq respectively), and 0.00005U or 0.0001U of RNase V1 (- or + Hfq respectively) was added. For RNA-IN 0.004 or 0.01 ng of RNase A (- or + Hfq respectively), 0.01U of RNase T1, and 0.00005U or 0.0001U of RNase V1 (- or + Hfq respectively) was added. G-lanes were produced by T1 digestion of RNA following the manufacturer’s directions (Ambion). RNase reactions proceeded for 15 minutes at 25°C before RNA was ethanol precipitated and resuspended in 7 µL formamide load dye (97.5% deionized formamide (v/v), 10 mM EDTA, 0.5x TBE, 3% xylene cyanol (w/v)) and ~2 µL was loaded onto a high resolution 10% polyacrylamide denaturing gel. Hydroxyl radical footprinting was performed as previously described (Jain & Tullius, 2008). Briefly, following binding reactions, 1 µL of freshly prepared H_{2}O_{2} (2.5% (v/v)), Fe(II)EDTA (32 mM ferrous ammonium sulphate, 88 mM EDTA; Bio Basic), and sodium ascorbate (60 mM; Bio Basic) was added to each 9 µL binding reaction. The final
concentration of hydroxyl radical reagents in each 12 µL reaction were: H₂O₂, 0.21% (v/v); Fe(II), 2.67 mM; EDTA, 7.33 mM; sodium ascorbate, 5 mM. Following incubation at 25°C for 10 min, samples were processed as described above for RNase treatment. For footprinting reactions, RNA-OUT was at a final concentration of 65 nM, RNA-IN at 45 nM, and Hfq at a final concentration of 99-4380 nM. Gels were dried and imaged with a phosphorimager (GE Healthcare). A sample of each binding reaction (RNA-IN and RNA-OUT, each Hfq concentration) was run on a 6% polyacrylamide gel to monitor complex formation under the conditions used (Figure S2.1). Quantitation of RNA-IN hydroxyl radical footprinting was performed using ImageQuant software.

2.4.6 Determination of RNA-OUT:RNA-IN pairing rates

RNA-IN:OUT pairing reactions were carried out by spotting 3.5 µL of 32P-labeled RNA-IN and RNA-OUT onto separate faces of an Eppendorf tube, mixing them with 24.5 µL of Hfq storage/binding buffer, and immediately removing 9 µL to separate tubes containing 1 µL of Hfq (or Hfq storage buffer), as indicated. Mixing was achieved by rapid pipetting. Final concentrations of reactants were: RNA-IN*, 0.85 nM; RNA-OUT*, 8.5 nM; Hfq, 45 nM. Incubation was at 37°C and, after the indicated times, each 10 µL reaction was added to tubes containing 30 µL H₂O and 40 µL phenol:chloroform:isoamyl alcohol. These were immediately vortexed (8 seconds) and centrifuged (3 second pulse-spin) before removing 10 µL of the aqueous phase and loading it directly on a 6% native polyacrylamide gel at 7 V/cm. After the last sample was loaded, electrophoresis was continued at 13 V/cm for 45 minutes. Gels were dried and imaged as described above. Bands representing RNA-OUT:IN paired species or free RNA-IN were quantified (ImageQuant). The percentage of paired species (relative to total RNA-IN counts) was plotted on the y-axis and time on the x-axis (Prizm). The resulting curves were fit by non-linear regression to the equation:

\[ A: B_t = A: B_{max} (1 - e^{-k_{obs} \times t}) \]

where A:Bₜ is the percentage of binary complex at time t and k_{obs} is the observed rate constant.
2.4.7 Mating out assay

Mating out experiments were carried out with DBH33 (hfq\textsuperscript{+}) and DBH16 (hfq\textsuperscript{−}) as donor strains and HB101 as the recipient strain. Plasmids encoding IS10-Kan (pDH602) and Hfq (pDH700, 701, 713 and pWKS30 as the ‘empty vector’ control) were co-transformed into donor strains and transformants were selected on M9-Glucose plates supplemented with arginine, kanamycin (50 µg/ml) and ampicillin (50 µg/ml). Donors and recipient strains were grown in liquid media as previously described in (Ross et al, 2010) and mating was allowed to proceed for 1 hour whereupon mating mixes were pelleted and re-suspended in 0.85% saline. Re-suspended mating mixes were then plated on M9 media supplemented with glucose, leucine and streptomycin (150 µg/ml) or streptomycin plus kanamycin (50 µg/ml). Plating on the former gave the mating frequency and plating on the latter gave the number of transposition events. Relative transposition frequencies were calculated by dividing the number of colonies present on ‘streptomycin/kanomycin plates’ by the number of colonies on ‘streptomycin plates’. For statistical analysis, we first carried out an F-test to demonstrate that the variances between the hfq\textsuperscript{+} control group and the other treatments were not equal (hfq\textsuperscript{+} vs hfq\textsuperscript{−}/hfq\textsubscript{WT}, P = 0.001; hfq\textsuperscript{+} vs all other treatments, P < 0.0001). We then conducted a two-tailed t test with Welch’s correction (does not assume equal variances) to compare the various treatments to the hfq\textsuperscript{+} control group. All statistical analyses were carried out in Prizm. Sample numbers and P values for the t-test are reported in Figure 2.9.
2.5 Supplemental Material

Figure S 2.1. RNA footprinting EMSAs.

Following binding reactions (see Materials and Methods) for RNA-OUT (A) and RNA-IN (B) footprinting experiments (shown in Figures 3 and 4 respectively), a sample was run on a 6% polyacrylamide gel to monitor RNA:Hfq complex formation. The final
concentration of RNA-OUT was 65 nM, RNA-IN was 45 nM, and Hfq concentrations are reported per hexamer.
Figure S 2.2. RNA-IN-160 RNase footprinting
5' $^{32}$P-IN-160 (65 nM) was incubated in the absence (-) or presence (+) of Hfq (365 nM) before digestion with RNase A, T1, or V1 (lanes 5-10) or hydroxyl radical (3,4). A G-lane (G, 1) and untreated RNA (UT, 2) are shown. Changes in reactivity in the presence of Hfq are indicated by upward- or downward-facing triangles (increased and decreased reactivity, respectively). Sensitivity to single-strand specific nuclease A/T1 are indicated in red, while sensitivity to double-strand specific V1 is indicated in blue.
Figure S 2.3. Structure of RNA-OUT and RNA-IN in the presence of Hfq.

5’ end-labeled RNA-OUT (A) and RNA-IN (B) was probed with RNase A/T1/V1 in the presence of Hfq (see Figures 3 and 4). Nucleotides indicated in red and blue represent mFold constraints for ssRNA and dsRNA, respectively. These structures formed at 37°C.
in the presence of 10 mM MgCl$_2$. Putative Hfq-binding sites were entered as single-stranded constraints.
Figure S 2.4. IC50 determinations for A18 and/or DsrA competition experiments.

The percentage inhibition of Hfq:IN-1 (A-C) and Hfq:OUT-1 (D-E) complex formation by competitor RNAs (DsrA, A18 or DsrA+A18) is plotted as a function of competitor concentration. The data is derived from experiments in Figures 6 and 7 plus additional experiments not shown. Data were fit to sigmoidal curves to obtain the IC_{50} values (reported in Table 2.2).
Figure S 2.5. Measurement of RNA-OUT:RNA-IN pairing as a function of Hfq concentration.

RNA-IN:RNA-OUT pairing reactions were performed in the presence of the indicated concentrations of Hfq (WT or K56A) as described in Figure 2.8 for a 2 minute reaction. As 45 nM gave an enhancement in RNA-IN:OUT pairing, this concentration of Hfq was used in the experiments summarized in Figure 2.8.
**Figure S 2.6. Hfq western blots.**

**(A)** Hfq expression in donor strains used in mating out assays. Just prior to mating, aliquots of donor cultures were pelleted and resuspended in denaturing SDS load mix (2% SDS, 10% glycerol, 50 mM Tris-HCl pH 6.8, 0.25% bromophenol blue and 0.8 M β-mercaptoethanol). After heating at 95°C for 5 minutes, cell lysates were subjected to SDS-PAGE on a 14% polyacrylamide gel. Proteins were transferred to a PVDF membrane (Roche) and Hfq was detected by Western blot with a polyclonal rabbit anti-Hfq antibody (Gift of G. Storz). The primary antibody was diluted to 1:4000 in TBST; the secondary antibody (anti-rabbit IgG-HRP conjugate; Promega) was used at 1:10,000. Hfq was visualized with a SuperSignal West Pico kit (Thermo Scientific) and an AlphaImager. The position of monomeric Hfq and species (X) that cross react with the antibody are shown. Samples analyzed are described in Figure 2.9.  

**(B)** SDS-PAGE to assess the purity of Hfq preparations. Purified Hfq at the indicated concentrations was mixed with an equal volume of denaturing SDS load mix, heated to 95°C for 5 minutes, and 20 µL were run on a 12% SDS polyacrylamide gel. The gel was stained with...
Coomassie R250 and destained with 10% Acetic acid/30% Methanol (v/v). The species corresponding to monomer, and a species we presume to be hexamer, are indicated to the right of the image. The Hfq in lanes 9-11, denoted “CP”, were purified by cobalt column chromatography and polyA resin as described in Materials and Methods.
Rifampicin-sensitive strains DBH116 (hfq\(^{+}\)) and DBH117 (hfq\(^{-}\)) were transformed with plasmid pDH502 (Ap\(^{R}\)Kan\(^{R}\), source of IS10-Kan; (Ross et al, 2010)). Cells were grown to mid-log phase in 20 mL LB supplemented with 25 µg/mL kanamycin. (A) Two ‘time 0’ samples (600 µL) were removed to tubes containing 300 µL ‘Stop solution’ (1.5% [w/v] SDS, 300 mM Sodium Acetate, 30 mM EDTA), boiled for 1 minute and stored on ice. Rifampicin was immediately added to the remaining cells (final concentration: 200 µg/mL), and 600 µL samples were removed at the indicated time and processed as above. After all samples were collected on ice, total RNA was extracted by the ‘hot phenol’ method, treated with Turbo DNase (Ambion) and ethanol precipitated. Samples were resuspended in nuclease-free water and quantified by spectroscopy. RNA-OUT levels were assessed by primer extension with \(^{32}\text{P}\)-labeled JR4; cDNA was fractionated by denaturing PAGE. OUT (+69)* indicates full-length primer extension product; OUT (primer)* indicates unextended primer. Note that the top panel is an over-exposure of the...
region of the gel where +69* migrated. (B) The area of the bands corresponding to full-length RNA-OUT and unextended primer were quantified (ImageQuant) and RNA-OUT (+69) was divided by unextended primer for each lane; these values were normalized to time 0 in the hfq + strain and plotted on the y-axis, with Time (after rifampicin addition) on the x-axis. To obtain half-lives, the resulting curves were fit by non-linear regression to the equation describing exponential decay (Prizm); note that the plateau was set to zero RNA-OUT remaining:

\[ P = \text{Span} \cdot e^{-k \cdot t} \]

Where P is the proportion of RNA-OUT remaining, Span is 1.0, t is Time (in minutes) and k is the rate of decay in units of min^{-1}. The half-life \( t_{1/2} \) is equal to 0.6932/k.
2.6 References


Jain C (1995) IS10 antisense control in vivo is affected by mutations throughout the region of complementarity between the interacting RNAs. *Journal of molecular biology* **246**: 585-594


Kittle JD, Simons RW, Lee J, Kleckner N (1989) Insertion sequence IS10 anti-sense pairing initiates by an interaction between the 5' end of the target RNA and a loop in the anti-sense RNA. *Journal of molecular biology* **210**: 561-572


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


Chapter 3

3 Hfq binds directly to the ribosome binding site of IS10 transposase mRNA to inhibit translation

3.1 Introduction

Hfq is an abundant RNA-binding protein that acts at the core of complex post-transcriptional regulatory networks in many bacteria and is critical for stress and virulence responses (Sobrero & Valverde, 2012; Storz et al, 2011; Vogel & Luisi, 2011). It is found in at least 50% of sequenced bacteria (Sun et al, 2002) and has been predicted to be involved in the regulation of 269 mRNAs in *Escherichia coli* and at least 20% of all genes in *Salmonella* Typhimurium (Ansong et al, 2009; Guisbert et al, 2007; Sittka et al, 2008). Hfq is important for the function of trans-encoded small regulatory RNAs (sRNAs) that base-pair with partially complementary mRNAs. Hfq binds sRNAs and their partner mRNAs and facilitates intermolecular base-pairing. This typically affects translation and/or transcript stability. Hfq contains three RNA-binding surfaces all of which play a role in promoting base-pairing between RNAs. The ‘top’ and ‘bottom’ of the toroidal-shaped Hfq homohexamer are termed the proximal and distal RNA binding surfaces, respectively. The proximal surface binds short U-rich sequences typically found in sRNAs while the distal surface binds longer ARN repeats (where A is an adenine, R is a purine, and N can be any nucleotide) typically found in mRNAs (Link et al, 2009; Mikulecky et al, 2004). The proximal surface is proposed to be critical for sRNA stability through interactions with the 3’poly(U) tract following a Rho-independent terminator (Ishikawa et al, 2012; Sauer & Weichenrieder, 2011). The third, less defined surface consists of the outer rim or lateral RNA binding surface. This surface connects the proximal and distal RNA-binding sites. The lateral surface is extremely basic in *E. coli* and may be important for binding internal U-rich sequences of sRNAs (Sauer et al, 2011).

---

One model for Hfq-catalyzed pairing predicts simultaneous binding of a cognate sRNA and mRNA pair via the proximal and distal surfaces, respectively. Hfq binding sites are often just outside of RNA pairing sequences so simultaneous binding would tether the RNAs to Hfq while keeping seed regions available for pairing (Panja & Woodson, 2012). The RNAs can then initiate pairing by interacting in either the lateral or proximal surfaces and the RNAs are released as pairing proceeds (Hopkins et al, 2011; Hwang et al, 2011; Panja et al, 2013).

In addition to a role in sRNA-based regulation, Hfq has been shown to directly affect translation. In the case of sdhC mRNA, the sRNA Spot42 recruits Hfq to an AU-rich region in the translation initiation region (TIR) to inhibit translation. As the Spot42 pairing region in sdhC is too far upstream of the TIR to influence translation, it was inferred that stable association of Hfq with sdhC was sufficient to compete with 30S ribosomal subunit binding (Desnoyers & Masse, 2012). In another example, Hfq was shown to bind to a translational enhancer in cirA mRNA and block translation. Interestingly, in this case translation repression was relieved by the upstream binding of an sRNA (RyhB) that caused restructuring of the mRNA within the 5’ untranslated region (5’UTR), which ultimately prevented Hfq binding (Salvail et al, 2013). Finally, evidence has been presented in two different organisms that Hfq autoregulates expression by binding its own TIR (Sobrero & Valverde, 2011; Vecerek et al, 2005). No sRNAs have been implicated in this autoregulatory loop, supporting the contention that Hfq can act directly to inhibit translation. In the above examples, Hfq binding to the TIR of an mRNA is the effector of translational control, in contrast to sRNA-dependent regulation where the stable sRNA-mRNA duplex is responsible for blocking ribosome binding. Unlike sRNA-dependent regulation, the role of each RNA-binding surface of Hfq in direct translational repression is largely unknown. However, Hfq binding to the TIR of target mRNAs would presumably require the distal surface, which preferentially binds purine rich sequences such as the Shine-Dalgarno sequence.

In addition to its role as an important regulator of endogenous gene expression, Hfq was recently found to suppress Tn10/IS10 transposition in E. coli (Ross et al, 2010). Tn10 is a composite transposon containing genes encoding for tetracycline resistance.
Its component insertion sequence IS10-Right encodes a functional transposase that catalyzes the chemical steps in Tn10/IS10 transposition (Chalmers et al, 2000; Foster et al, 1981; Halling et al, 1982). Expression of IS10 transposase is regulated by Dam methylation as well as a 69 nt antisense RNA (asRNA) that is transcribed from the opposite strand of DNA relative to the transposase (Roberts et al, 1985; Simons & Kleckner, 1983). The first 35nt of this asRNA (RNA-OUT) is perfectly complementary to the TIR of the transposase mRNA (RNA-IN), and pairing of these two RNAs inhibits translation by preventing ribosome binding (Figure 3.1A) (Ma & Simons, 1990).

Antisense control of transposase expression increases with IS10 copy-number, a phenomenon termed ‘multi-copy inhibition’ (MCI). MCI can be explained by the fact that transposase is a cis acting protein whereas the asRNA is trans acting (Jain & Kleckner, 1993). Accordingly, increasing transposon copy-number essentially serves to increase the amount of trans acting inhibitor while the effective concentration of transposase per element remains constant. Importantly, a single-copy IS10 element is not subject to antisense control of transposase expression (Kleckner, 1990).

Hfq was initially linked to Tn10/IS10 transposition when it was found that IS10 transposition increased in the order of 80-fold in an hfq strain of E. coli harbouring IS10 on a multi-copy plasmid. In contrast, the impact of Hfq-deficiency on transposition was greatly reduced (7-fold increase), but not completely abrogated, when transposition was measured for IS10 in single copy. These observations were consistent with Hfq contributing to MCI, but also playing a role in down-regulating IS10 transposition independent of the MCI pathway (Ross et al, 2010). Subsequent work demonstrated that Hfq bound both RNA-IN and RNA-OUT in vitro, and accelerated the rate of IN-OUT pairing almost 20-fold, observations that are consistent with Hfq working through its prototypical RNA pairing pathway to promote the MCI response (Ross et al, 2013). It remains to be determined how Hfq regulates Tn10/IS10 transposition when MCI is not in play. However, it has been shown that: (i) in the absence of RNA-OUT transposase expression increased 6-fold in hfq, (ii) there is an Hfq binding site in RNA-IN that overlaps the TIR and (iii) Hfq status has only a subtle effect on steady-state RNA-IN
Figure 3.1. Overview of the Tn10/IS10 system.

(A) The structure of Tn10 is shown (Chalmers et al, 2000). IS10R encodes a functional transposase protein that catalyzes the chemical steps in Tn10/IS10 transposition. In addition to transposase mRNA (RNA-IN, blue), IS10 encodes an asRNA (RNA-OUT, red) that represses transposase translation by blocking ribosome binding. Hfq represses transposase translation by facilitating antisense pairing as well as through an antisense-independent mechanism. OE and IE are outside and inside ends respectively. (B) Schematic of the three IS10R constructs used in this work. The promoters for RNA-IN (pIN) and RNA-OUT (pOUT) are indicated with blue and red boxes respectively and the transcriptional start sites are shown. RNA-OUT terminates at nucleotide 47 of IS10, which is indicated with a dashed line. The DNA sequence of pIN is shown with the location of two nucleotide changes (R5 and HH104) that each destabilize RNA-OUT. The DNA adenine methyltransferase (DAM) site, which overlaps with the -10 region, is
also shown. The IS10\textsuperscript{HH104}-kan construct consists of a kanamycin resistance gene inserted downstream of the transposase stop codon but upstream of the IE and both translational dusions consist of the indicated portion of IS10R fused to codon 10 of \textit{lacZ}. In A and B, black arrows indicate the polarity of each open reading frame (ORF).
transcript level (Ross et al, 2013; Ross et al, 2010). Taken together it seems likely that Hfq functions in the MCI-independent pathway by inhibiting translation of RNA-IN. The goal of the current work was to test this hypothesis and to further characterize how Hfq interacts with RNA-IN.

3.2 Results

3.2.1 Antisense-independent regulation of transposase expression requires the distal surface of Hfq

To gain further insight into how Hfq regulates IS10 transposase expression (independent of its cis-encoded sRNA) we asked if regulation was maintained when RNA-binding surfaces of Hfq were mutated. Our expectation was that if a trans-encoded sRNA is involved in this pathway, all three surfaces would be critical for regulation. In particular, we expected the proximal surface to be important for stabilizing any involved sRNAs and the lateral surface for catalyzing pairing with RNA-IN.

To assess the function of each Hfq binding surface in antisense-independent regulation of transposase expression, we assayed the ability of wild-type and mutant forms of Hfq to complement an hfq− phenotype. We constructed a chromosomal IS10−339-lacZ translational fusion with a single bp change (HH104) in the promoter for RNA-IN which increases transcription ~100-fold (Figure 3.1B) (Case et al, 1988). The HH104 mutation also destabilizes RNA-OUT; however, a single-copy transposase-lacZ translational fusion would not normally be regulated by the cis-encoded RNA-OUT so the net effect of this mutation is to simply increase RNA-IN expression to detectable levels (Case et al, 1989). Expression of transposase-lacZ was measured in an hfq− strain of E. coli harbouring plasmids expressing HfqWT or Hfq deficient in RNA-binding at the distal (Y25A), proximal (K56A), or lateral (R17A) surface (Mikulecky et al, 2004; Panja et al, 2013).

We show in Figure 3.2A that transposase-lacZ expression increased almost 13-fold in the absence of Hfq, and that regulation was fully restored when HfqWT was
Figure 3.2. Impact of mutant forms of Hfq on IS10 transposase expression and transposition.

(A) Transposase expression was measured in the context of a chromosomal transposase-lacZ translational fusion (parent strain DBH298) with the indicated forms of Hfq expressed or in the absence of Hfq expression. The bars show β-galactosidase activity (Miller units) with standard error of the mean, measured in mid-exponential phase in LB (n=8). Where indicated, the hfq- strain (DBH299) was transformed with a low-copy plasmid encoding Hfq from its native promoter (P3). The mean relative expression for each strain is indicated at the top of the graph, where transposase-lacZ in hfq+ was set at 1. (B) Transposition of a chromosomal IS10HH104-kan was measured by the conjugal mating out assay (see Materials and Methods) in an hfq- strain (DBH337) transformed with one of the indicated Hfq-encoding plasmids. The mean relative transposition frequency for each strain is indicated at the top of the graph, where transposition in the presence of HfqWT was set at 1. Error bars indicate the standard error of the mean for two independent experiments (n=8).
expressed from a plasmid. In contrast, none of the Hfq variants were able to fully complement \( hfq^- \), with \( Hfq^{Y25A} \) being the most impaired and \( Hfq^{R17A} \) functioning the most like WT. Importantly, the reduced function of the mutant proteins cannot be attributed to protein expression as the Hfq levels were not significantly different for WT versus the mutant forms of Hfq (Figure S3.1). As the results showed that the integrity of the lateral site is not important for regulation and the integrity of the proximal site is less important than that of the distal site, our data suggested that Hfq repression was sRNA-independent.

We also performed the Hfq complementation experiment in the context of an IS10 transposition assay. As IS10 transposition frequency is directly proportional to transposase expression (Morisato et al, 1983), this experiment allowed us to indirectly measure the effect of Hfq mutations on native transposase expression. Transposition of a single-copy IS10\(^{HH104}\)-kan element (Figure 3.1B) in an \( hfq^- \) strain of \( E. coli \) was repressed 13-fold in the presence of \( Hfq^{WT} \) and close to full repression was achieved in the presence of \( Hfq^{K56A} \) and \( Hfq^{R17A} \) (Figure 3.2B). However, in accordance with the expression data (Figure 3.2A), \( Hfq^{Y25A} \) was the least effective of the mutant Hfq forms in repressing transposition.

It is not clear why the \( Hfq^{K56A} \) functioned essentially as WT in the transposition assay (Figure 3.2B) while exhibiting a moderate defect in repressing transposase-lacZ expression (Figure 3.2A). However, the concordance of the other Hfq variants between the two experiments lead us to conclude that the distal surface is critical for repressing transposase expression in the absence of RNA-OUT, while the proximal and lateral surfaces are dispensable for regulation.

### 3.2.2 The distal surface of Hfq binds the RBS of RNA-IN \textit{in vitro}

We have previously shown that Hfq binds a 14nt A-rich region of RNA-IN which overlaps with the ribosome binding site (RBS), as well as an 8nt U-rich region overlapping with codons 5-8 of the transposase coding region (Ross et al, 2013). Based on the results from the previous section, we anticipated that the A-rich binding site within the TIR would be the critical site for Hfq regulation as it has the signature of a distal binding site. To further characterize the two Hfq binding sites in the 5’ segment of RNA-
IN, we carried out chemical and enzymatic RNA footprinting. Purified WT, Y25A, and K56A Hfq were incubated with an RNA corresponding to the first 160 nt of RNA-IN (IN-160) followed by partial digest with lead acetate (Pb$^{2+}$), RNase T1, or RNase V1.

Within the first 70 nt of RNA-IN three regions of cleavage reagent protection were detected in the presence of Hfq$^{WT}$ (Figure 3.3). The most upstream region (relative to the start codon) spans the ribosome binding site (RBS), extending roughly from nt -20 to -4; hereafter this site will be referred to as site 1. Additional regions of protection downstream of this include residues 13-20 and 32-34; hereafter referred to as sites 2 and 3, respectively. Another region spanning nt 3-7 became hypersensitive to RNase T1 and showed reduced RNase V1 cleavage, consistent with Hfq binding inducing a structural transition in this region from dsRNA to ssRNA. In contrast, in reactions with Hfq$^{Y25A}$, site 1 showed greatly reduced protection from both V1 and lead cleavage, consistent with the distal surface of Hfq making contacts with this site. The V1 and lead cleavage pattern in site 1 for Hfq$^{K56A}$ was very similar to that of Hfq$^{WT}$. As Hfq$^{K56A}$ retains a fully functional distal binding surface, this result further supports our conclusion that the distal surface of Hfq contacts site 1.

At sites 2 and 3 the K56A mutation greatly reduced protection from both V1 and lead cleavage whereas the Y25A mutation did not. This is consistent with the proximal surface of Hfq binding both sites 2 and 3.

Finally, the T1 hypersensitivity at positions 3-7 was lost only in the reaction with Hfq$^{Y25A}$. Accordingly, we infer that the structural transition in this region is dependent on the distal face of Hfq binding to site 1.
Figure 3.3. RNA-IN footprinting with WT and mutant forms of Hfq.
5\textsuperscript{32}P-labeled RNA-IN (40 nM) was incubated with his-tagged WT, Y25A or K56A Hfq (750 nM Hfq hexamer) before treatment with Pb\textsuperscript{2+}, RNase T1, or RNase V1. Untreated RNA was also incubated with or without WT Hfq (lanes 1-2). Nucleotide numbering is relative to the translational start codon (AUG), where the A is position 1. Positions that were protected from cleavage by WT and K56A but not Y25A are indicated with red arrows, while positions protected by WT and Y25A but not K56A are indicated with blue arrows. RNase T1 hypersensitivities are highlighted with black arrows. The nucleotide sequence of the first 160nt of RNA-IN is shown below the gel image with distal and proximal specific-binding sites indicated in red and blue respectively. The RBS and start codon are underlined.
3.2.3 Hfq binding to site 1 is critical for regulation of transposase expression *in vivo*

As the distal surface of Hfq is most critical for repression of transposase expression and transposition, our footprinting data suggests that Hfq primarily exerts its regulatory role by binding site 1, which overlaps the RBS of RNA-IN. We set out to further test this hypothesis by analyzing the impact of nucleotide changes in site 1 and site 2 on Hfq interactions and ultimately transposase expression.

Our first objective was to define the relative binding affinity of sites 1 and 2 for Hfq. We show by EMSA in Figure 3.4 that in the presence of non-specific competitor RNA, Hfq\(^{WT}\) forms two distinct complexes with RNA-IN-160, which we term IN:Hfq-1 and IN:Hfq-2. IN:Hfq-1 was detected at the lowest Hfq concentrations and appears to be converted into IN:Hfq-2 as Hfq concentrations increased. Apparent K\(_D\) values were calculated to be 0.2 nM for IN:Hfq-1 and 46.2 nM for IN:Hfq-2. Multiple mutations in site 1 were required to reduce formation of IN:Hfq-1. For example, the M5 mutant contains 8 nucleotide changes in site 1 and increased the K\(_D\) for IN:Hfq-1 by just over 30-fold and essentially abrogated the formation of IN:Hfq-2. In contrast, the site 2 mutant M2, which contains 5 nucleotide changes, had little impact on IN:Hfq-1 formation, but increased the K\(_D\) for IN:Hfq-2 almost 4-fold. Note that because of the manner in which these mutants were identified we do not know if all the nucleotide changes are necessary for the observed effects (see Materials and Methods).

Based on these results, as well as footprinting experiments (Figure S3.2), we conclude that IN:Hfq-1 is formed through Hfq binding to site 1 and IN:Hfq-2 is formed through Hfq binding to both site 1 and site 2; note we have not looked at the importance of site 3 in IN:Hfq complex formation. Moreover, Hfq binds site 1 with a much higher affinity than it binds site 2 and Hfq binding to site 2 appears to be dependent on Hfq first binding site 1. Since formation of IN:Hfq-2 increased in a concentration-dependent manner, it is likely that occupancy of site 2 depends on recruitment of a second Hfq hexamer to IN:Hfq-1, as opposed to the unoccupied proximal surface of an Hfq hexamer bound at site 1 engaging site 2.
Figure 3.4. Effect of RNA-IN mutations on Hfq binding.

(A) Sequence of the first 50nt of RNA-IN with the RBS and start codon indicated in blue and the sequence of two Hfq-binding sites (site 1 and site 2) underlined. Nucleotide changes for M5 and M2 mutants are shown. (B) Hfq binding to 32P-labeled INWT-160, INM5-160, and INM2-160 was measured by EMSA. Binding reactions contained 20 ng/µl of total yeast RNA, the indicated concentrations of Hfq (reported per hexamer) and ~1 nM RNA-IN. Band intensities of the representative gel images shown were quantified and the percent complex formed was plotted against Hfq concentration to calculate apparent dissociation constants (K_D). Hfq-RNA-IN interactions for IN:Hfq-1 and IN:Hfq-2 are described by K_D1 and K_D2 respectively.
We next looked in vivo at the impact of disrupting Hfq interactions with sites 1 and 2. To do this we introduced the M5 and M2 mutations into an IS10\textsubscript{1,242}-lacZ translational fusion on a multi-copy plasmid. A multi-copy transposase-lacZ translational fusion would normally be highly repressed by the cis-encoded RNA-OUT. To separate the role of Hfq in direct repression of transposase expression from its role in facilitating antisense pairing, we introduced a single nucleotide mutation (R5) into the promoter region of RNA-IN which destabilizes RNA-OUT while having only a subtle effect on RNA-IN transcription (Figure 3.1B) (Case et al, 1988). Based on our earlier results (Figure 3.2A,B), our expectation was that the M5 mutation would make IS10-lacZ expression insensitive to Hfq status. As the two mutants have multiple nucleotide substitutions, we were concerned that these changes could have indirect effects on transposase expression. Accordingly, we also isolated RNA from cells used in the reporter assays and performed primer extension analysis to monitor steady-state transcript levels (Figure 3.5).

We show in Figure 3.5 that both M5 and M2 mutants exhibited reduced Hfq regulation, with the degree of dysregulation being stronger for M5 versus M2. In this reporter set-up there was a 3.2-fold decrease in transposase expression in the presence versus the absence of Hfq, consistent with Hfq having a negative regulatory role. In contrast, for the M5 reporter, expression levels were essentially the same in hfq\textsuperscript{+} and hfq and for the M2 reporter, expression decreased about 1.5-fold in the presence of Hfq.

In the above experiment the ratio of transposase-lacZ expression (β-galactosidase assay) to the steady state level of fusion transcript provides a measure of the translation efficiency. For example, if expression was low and transcript levels were high, this would be indicative of low translation efficiency. The presence of Hfq in cells expressing the WT reporter decreased translation efficiency approximately 2.5-fold, consistent with Hfq interfering with transposase translation. For the M5 reporter, translation efficiency was greatly reduced compared to the WT reporter, which is not unexpected given that 5 of the 8 nt changes in this construct are in the RBS. Importantly, the translation efficiency did not further decrease in the presence of Hfq. Thus Hfq is unable to down-regulate translation when site 1 is mutated. Translation efficiency for the M2 reporter was
Figure 3.5. Impact of mutant forms of RNA-IN on transposase expression in
\( hfq^+ / hfq^- \) strains.

Plasmids encoding WT and mutant forms of a transposase-\( \text{lacZ} \) translational fusion were transformed into \( hfq^+ \) (DBH107) or \( hfq^- \) (DBH12) cells and after growth of transformants to mid-exponential phase in LB media, \( \beta \)-galactosidase activity was measured. Error bars show the standard error of the mean for three independent experiments (\( n=12 \)). RNA was extracted from cells immediately before the Miller assay and RNA-IN was detected by primer extension (lower panel). \( lpp \) mRNA was used as a loading control. The relative transcript level from two isolates of each strain was quantified and normalized to WT RNA-IN in \( hfq^+ \) (set at 1). ITR (lane 1) refers to \textit{in vitro} transcribed RNA-IN and served as a positive control for primer extension, and ‘-‘ (lane 2) is RNA from cells without a IS10-\( \text{lacZ} \) plasmid. The relative translation efficiency for each strain was calculated by dividing Miller units by relative transcript levels (shown as circles on the graph).
intermediate to that of the WT and M5 reporters indicative of site 2 playing a more minor role compared to site 1 in Hfq-directed repression of translation.

### 3.2.4 Hfq blocks 30S ribosome binding to RNA-IN *in vitro*

Our results thus far show that Hfq inhibits IS10 transposase expression *in vivo*, and that the most important Hfq-RNA interaction for this response is between the distal surface of Hfq and site 1, which includes the RBS of RNA-IN. In addition, results from Figure 3.5 are consistent with Hfq down-regulating IS10 transposase expression by interfering with IS10 transposase translation. To further test the hypothesis that Hfq binding to site 1 inhibits RNA-IN translation, we performed *in vitro* toeprinting assays. We show in Figure 3.6A (lane 6) that addition of the 30S ribosomal subunit and initiator tRNA to RNA-IN resulted in a strong block of reverse transcription at position +16 relative to the RNA-IN start codon, with minor pauses at nts +17/+18 as has previously been reported (Ma & Simons, 1990). These observations are consistent with a stable translation initiation complex forming on RNA-IN. When Hfq was added prior to addition of the 30S ribosome and initiator tRNA (lanes 7-12), there was a decrease in the toeprint signal, the magnitude of which was dependent on the Hfq concentration. For example, at an Hfq concentration of 200 nM, where Hfq and RNA-IN are present at a 1:1 molar ratio, the toeprint signal decreased greater than 90% relative to the signal observed in the absence of Hfq (Figure 3.6A,B). In contrast, when the same experiment was performed with *lpp* or *usg* mRNA, inhibition of the toeprint signal was significantly weaker (Figure S3.3A). For example, at the same ratio of Hfq:mRNA that gave greater than 90% inhibition for IS10, only 50% inhibition was observed for *lpp* and *usg* (Figure 3.6B). Although Hfq plays a role in repressing *lpp* expression by stabilizing the sRNA MicL (Guo et al, 2014), there is no evidence that Hfq directly interacts with *lpp* mRNA *in vivo* (Bilusic et al, 2014; Chao et al, 2012) (see also Figure 3.9). Similarly, Hfq does not interact with *usg* *in vivo* (Beisel et al, 2012). Accordingly, the toeprinting results in Figure 3.6A are consistent with Hfq acting specifically to block translation initiation in the IS10 system. We presume that the relatively low level of toeprint inhibition observed in the *lpp* and *usg* experiments is the result of non-specific interactions between Hfq and components of the 30S ribosome and thus represents ‘background noise’ in the assay.
Figure 3.6. Impact of Hfq on initiation of RNA-IN translation in vivo.

(A) 30S ribosome binding to RNA-IN +/- Hfq is shown in a toeprint assay (for details see Materials and Methods). Addition of 30S ribosomal subunits and initiator tRNA is indicated by +. The toeprint signal is indicated (+16/+17/+18) with numbering relative to the translational start codon. CUAG refers to sequencing reactions generated from the same RNA used for toeprinting. (B) Quantitation of RNA-IN (A), lpp, and usg (Figure S.3.3) toeprints. Toeprint signal was normalized to the combined band intensity (+16/+17/+18 for RNA-IN; +15/+16 for lpp and usg) in the absence of Hfq, which was set at 100. The dashed line highlights 50% inhibition of the toeprint signal. (C) Toeprint analysis of RNA-IN with distal (A18) and proximal (U7) site-specific competitor RNAs. In addition, RNA-IN toeprint was also analyzed for reactions containing Y25A versus WT Hfq. Note that for this latter comparison, Hfq contained a C-terminal 6x-His epitope tag. The competitor RNAs alone had no effect on ribosome binding (not shown). (D)
Band intensities in C were quantified and normalized to the toeprint signal in lane 5 where no competitor or Hfq was added.
We reasoned that since the distal surface of Hfq interacts with site 1 of RNA-IN, then the distal surface of Hfq would be critical for blocking 30S ribosome binding to RNA-IN. Accordingly, a competitor RNA that is specific for the distal RNA binding face of Hfq should relieve the ‘toeprint repression’ afforded by Hfq. We show in Figure 3.6C and D that when Hfq was pre-incubated with A18, a distal site binding RNA, prior to its addition to RNA-IN the toeprint signal increased approximately 2.5-fold relative to a control reaction where no competitor was added (compare lanes 6 and 7). Also, if a proximal face binding RNA (U7) was used instead of A18, there was no increase in the toeprint signal (compare lanes 6 and 8). We also show in Figure 3.6C and D that Hfq^{Y25A} failed to reduce the toeprint signal (compare lanes 9 and 10 with lanes 11 and 12). Taken together, the results in this section show that Hfq can inhibit 30S ribosome binding to RNA-IN in vitro and that this inhibition requires an available distal surface on Hfq. Furthermore, as we have shown that there is a high affinity Hfq binding site (site 1) that spans the RBS of RNA-IN and engages the distal binding surface of Hfq, we conclude that Hfq binding to this site is responsible for blocking translation initiation.

3.2.5 Hfq binds native RNA-IN in vivo and this binding is inhibited by site 1 mutations

As Hfq is a pleiotropic regulator of gene expression in *E. coli* there is a concern that any phenotype observed in an *hfq* strain might be the result of dysregulation of a factor under Hfq control. As we are proposing that Hfq binds directly to the 5’UTR of RNA-IN to block translation, we thought it important to look for Hfq-RNA-IN binding in vivo where RNA-IN would have to compete with other cellular RNAs for Hfq binding. We performed an Hfq-RNA immunoprecipitation experiment (RIP) with *hfq* cells containing a chromosomal copy of IS1^H104-kan and a plasmid encoding FLAG-tagged Hfq (Figure 3.7A). Importantly, this experiment used the same full-length RNA-IN as that used in our transposition experiments.

RNA recovered from an Hfq IP was subject to RT-PCR where RNA-IN was amplified from total cDNA with gene specific primers (see Materials and Methods and Figure S3.4). PCR reactions were then analyzed on an agarose gel. We show that RNA-IN was strongly enriched in the Hfq IP compared to control reactions (Figure 3.7A).
Figure 3.7. Hfq-RNA-IN immunoprecipitation (RIP) assay.

(A) *hfq* cells containing a chromosomal IS10\(^{HH104}\)-kan element (DBH337) were transformed with plasmids expressing Hfq\(^{WT}\) (pDH904) or Hfq\(^{WT}\)-3xFLAG (pDH909; C-terminal 3xFLAG tagged Hfq). Hfq was immunoprecipitated (IP) from cell lysates with ANTI-FLAG\(^{TM}\) M2 magnetic beads; untagged Hfq (Hfq\(^{WT}\)) served as a negative control. The first 160 nt of RNA-IN (top panel) or nts 1071-1425 of 16S rRNA were detected by RT-PCR. Samples were analyzed on a 2% agarose gel that was stained with ethidium bromide. No reverse transcription (-RT) controls are shown (lanes 2, 4, 6, and 9). L is a DNA ladder (lane 1). (B) *hfq* cells (DBH337) were co-transformed with Hfq\(^{WT}\)-3xFLAG plasmid and a plasmid encoding either WT or M5 transposase-*lacZ*. Hfq RIPs were performed as in A. Band intensities for the input and IP RT-PCR signal were quantified with an AlphaImager 3400 (Alpha Innotech).
Specifically, when cells expressing an untagged Hfq (Hfq\textsuperscript{WT}) were subject to RIP, no Hfq was detected in the IP fraction (Figure S3.4A) and accordingly no RNA-IN was detected by RT-PCR (Figure 3.7A). We therefore conclude that RNA-IN is a \textit{bona fide} Hfq binding partner \textit{in vivo}.

We also performed an Hfq IP with cells containing either the WT or M5 IS10-lacZ translational fusion on a multi-copy plasmid. RT-PCR analysis of these samples revealed a 2.3-fold reduction in the recovery of IN\textsuperscript{M5}-lacZ compared to IN\textsuperscript{WT}-lacZ in the IP. This result is consistent with the Hfq binding site within the 5’UTR of RNA-IN (site 1) providing important determinants for Hfq binding \textit{in vivo} (Figure 3.7B).

### 3.2.6 ChiX overexpression titrates Hfq away from RNA-IN

Overexpression of Hfq-binding RNAs can impinge on other regulatory networks by titrating available Hfq away from other mRNAs and sRNAs (Hussein & Lim, 2011; Moon & Gottesman, 2011; Papenfort et al, 2009). Given our finding that RNA-IN can compete with other cellular RNAs for Hfq binding (Figure 3.7), we wondered if induction of an sRNA might increase transposase expression by sequestering Hfq away from RNA-IN. Although most sRNAs interact with the proximal surface of Hfq, an overexpressed sRNA that can bind the distal surface of Hfq might block Hfq’s association with the RBS of RNA-IN and therefore increase transposase translation by allowing the ribosome to bind. To test this prediction we screened a library of Hfq-binding sRNAs to see if any increased transposase-lacZ expression.

We transformed \textit{hfq}\textsuperscript{+} cells containing the chromosomal IS10\textsubscript{1,339}-lacZ translational fusion with a plasmid expressing one of 14 sRNAs (Sgrs, ChiX, RybB, FnrS, MicC, RydC, MgrR, RprA, RyeB, CyaR, MicF, GlmY, MicA, and GcvB) or a vector control (Mandin & Gottesman, 2010). Transformants were subject to blue-white screening on X-gal plates. Our screen identified a single sRNA, ChiX, which increased IS10 transposase expression. Notably, the distal surface of Hfq has been previously shown to be important for ChiX stability and Hfq binding \textit{in vivo} (Zhang et al, 2013).
We next quantified the level of IS10 transposase up-regulation by measuring β-galactosidase activity with overexpression of ChiX. As a control, we included SgrS, as this sRNA did not give a blue colony color in our screen. As shown in Figure 3.8A, overexpression of ChiX increased transposase-lacZ expression almost 12-fold compared to a vector control, while SgrS overexpression had no effect on transposase expression. We also analyzed RNA extracted immediately before the Miller assay by Northern blot and primer extension to measure sRNA induction and transposase-lacZ transcript levels, respectively (Figure 3.8B,C). Consistent with previous results, a low amount of endogenous ChiX was detected in all samples (Figueroa-Bossi et al, 2009; Vogel et al, 2003), but there was a large increase in the presence of the ChiX plasmid (52-fold). This induction is comparable to that seen as cells transition to stationary phase (Vogel et al, 2003). Importantly, ChiX induction resulted in only a 2-fold increase in transposase-lacZ transcript levels relative to the vector control (Figure 3.8C). As ChiX overexpression increased transposase expression 12-fold while having only a subtle effect on steady-state transcript levels, we conclude that ChiX increases transposase translation and not transcription or mRNA stability.

ChiX may increase transposase translation by one of two mechanisms: (i) ChiX may base-pair with RNA-IN to increase ribosome accessibility, as seen in the rpoS system (Brown & Elliott, 1997; Soper et al, 2010), or (ii) ChiX may bind Hfq with high affinity and block Hfq-binding to RNA-IN. We used in vitro lead footprinting with 5’labeled ChiX and purified Hfq to define Hfq binding sites on ChiX. We also included RNA-IN in the footprinting reactions to determine if ChiX base-pairs with RNA-IN.

In the absence of Hfq, ChiX exhibited high reactivity to lead with the exception of nucleotides 17-22, 56-65, and 70-78, which is consistent with a mostly unstructured RNA containing a 5’stem-loop and a Rho-independent terminator (Figure 3.8D, lane 5 and right panel). In the presence of Hfq, two regions of reduced lead cleavage consisting of nucleotides 24-33 and 48-52 were observed (Hfq-I and Hfq-II, compare lanes 5 and 6). Binding reactions were also analyzed by EMSA (shown beneath footprinting gel image) and showed that Hfq forms a single complex with ChiX sRNA (ChiX:Hfq, lane 2 and 3).
Figure 3.8. ChiX positively regulates IS10 transposase translation.

(A) *hfq*+ containing a chromosomal IS101-339-lacZ translational fusion were transformed with a plasmid expressing ChiX (pDH765), SgrS (pDH764), or a vector control (pDH763). Transformants were grown to mid-exponential phase in LB media and β-galactosidase activity was measured. Error bars show the standard error of the mean for two independent experiments (n=7) and the relative expression is shown above the graph, where transposase-lacZ in the presence of vector was set to 1. RNA was extracted...
immediately before the Miller assay. (B) 3.5 µg of total RNA (three biological isolates) was used for a Northern blot using a 5′[^32]P-labeled oligonucleotide (SgrS) or internally labeled antisense RNA probe (ChiX, 5S rRNA). (C) Primer extension analysis of 10 µg of total RNA (four biological isolates) was used to detect RNA-IN-lacZ transcript and *lpp* was analyzed as an internal control. The ratio of IN-lacZ:*lpp* was normalized to the vector control and is shown with standard error of the mean as a graph above the gel images. (D) 5′[^32]P-labeled ChiX RNA (100 nM) was incubated with purified Hfq or Hfq and *in vitro* transcribed RNA-IN before limited cleavage with Pb[^2+]. An RNase T1 sequencing lane (G; lane 1) and untreated controls (lanes 2-4) are shown. An aliquot of binding reactions was analyzed on a 6% native polyacrylamide gel (bottom panel). The secondary structure of ChiX is shown (right panel) with three Hfq-binding sites (Hfq-I/II/III) highlighted in red.
We also performed footprinting experiments with Hfq binding face mutants and ChiX RNA (Figure S3.5). We show that the distal surface of Hfq binds Hfq-I, while binding of Hfq-II requires an intact proximal surface. In addition, a third Hfq binding site (Hfq-III) was identified that includes the poly(U) tract following the Rho-independent terminator, which accordingly interacts with the proximal surface of Hfq.

In the presence of Hfq and a 5- or 7-fold molar excess of RNA-IN to ChiX there were no additional regions protected from lead cleavage (compare lane 5 to 7 and 8). This indicates that ChiX does not base-pair with the first 160nt of RNA-IN, a conclusion that is also supported by the absence of an additional complex (i.e. ChiX:RNA-IN binary complex or ChiX:Hfq:RNA-IN ternary complex) in the EMSA (bottom panel). Addition of RNA-IN did however reduce the lead footprint in the A-rich Hfq binding site of ChiX (compare lane 6 with lanes 7-8) and the amount of ChiX:Hfq complex formed (compare lanes 3 and 4 in the EMSA), consistent with ChiX and RNA-IN competing for Hfq binding.

Based on our in vitro data that RNA-IN can compete with ChiX for Hfq binding we performed an RIP in hfq cells containing the IS101-339-lacZ translational fusion, FLAG-tagged Hfq, and the ChiX overexpression plasmid or a vector control. RNA recovered from the Hfq IP was analyzed directly by northern blot or primer extension to detect ChiX and IN-lacZ respectively; the 5S rRNA and lpp were also analyzed as negative controls (Figure 3.9). Overexpression of ChiX resulted in a 4.5-fold reduction in the amount of RNA-IN associated with Hfq. This experiment also allowed us to compare the relative binding affinities of ChiX and RNA-IN for Hfq. In the presence of the vector control, ChiX binds Hfq about 30-fold better than RNA-IN in vivo. Based on the differences in the amount of RNA analyzed (10 µg Input RNA, 0.3 µg IP RNA) we calculated that ChiX was enriched 365-fold in the Hfq IP, while RNA-IN was enriched 12-fold. We presume that the relatively low amount of 5S rRNA and lpp mRNA that were detected in the IP represent non-specific interactions with Hfq in vivo or during the IP procedure.
Figure 3.9. ChiX competes with RNA-IN for Hfq-binding *in vivo*.

*hfq* cells containing the chromosomal IS101-339-*lacZ* translational fusion (DBH299) were co-transformed with a plasmid encoding Hfq<sup>WT</sup>-3xFLAG (pDH909) and a plasmid expressing ChiX (pDH765) or vector control (pDH763). Hfq was immunoprecipitated from cell lysates with ANTI-FLAG<sup>®</sup> M2 magnetic beads. Total input RNA (10 µg) or RNA recovered from the IP (0.3 µg) was analyzed by Northern blot for ChiX and 5S rRNA, or primer extension for RNA-IN and *lpp*. Band intensities were quantified using ImageQuant.

<table>
<thead>
<tr>
<th>Vector</th>
<th>ChiX</th>
<th>IP/Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>IP</td>
<td>Input</td>
</tr>
<tr>
<td>ChiX</td>
<td>~10.95</td>
<td>10.50</td>
</tr>
<tr>
<td>5S rRNA</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>RNA-IN-<em>lacZ</em></td>
<td>0.36</td>
<td>0.08</td>
</tr>
<tr>
<td><em>lpp</em></td>
<td>0.05</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Since ChiX was the only sRNA in our screen that titrated Hfq away from RNA-IN, and ChiX is unique amongst sRNAs in that it contains a distal Hfq binding site, we wondered if other RNAs that interact with the distal surface of Hfq would increase transposase expression through an Hfq-titration mechanism. Most Hfq-binding mRNAs (including sodB, ptsG, and maeA) interact solely with the distal surface of Hfq (Zhang et al, 2013). We overexpressed the first 300nt of sodB, ptsG, or maeA mRNA and measured the impact on transposase expression. Note that the mRNAs were expressed from the same plasmid background as the sRNA overexpression library. Unexpectedly, transposase expression was mostly unaffected by mRNA overexpression (2.4-fold increase for sodB, 1.3-fold increase for ptsG and maeA) (Figure S3.6).

Together, the above results are consistent with ChiX activating IS10 transposase translation by titrating Hfq away from RNA-IN. ChiX does not interact with RNA-IN but does bind Hfq with high affinity and specificity and can compete with RNA-IN for Hfq binding in vitro and in vivo. Since overexpression of mRNAs containing a distal-binding site did not affect transposase expression, we think it likely that the ability of ChiX to up-regulate transposase expression is due to the fact that this sRNA possesses both distal and proximal Hfq binding sites.

3.3 Discussion

3.3.1 Direct repression of IS10 translation by Hfq

Hfq typically regulates translation by catalyzing pairing of an sRNA to the TIR of an mRNA. In the simplest model, Hfq simultaneously binds an sRNA and cognate mRNA near or overlapping the pairing sequences (also known as seed regions), and as pairing proceeds, Hfq is released from the sRNA-mRNA duplex, whereupon it can catalyze additional pairing reactions (Fender et al, 2010; Hwang et al, 2011; Panja & Woodson, 2012; Tree et al, 2014). The role of each Hfq binding surface in sRNA-dependent regulation has been studied extensively. A study of 7 different sRNA-mRNA pairs found that the proximal surface (in particular lysine 56) is critical for Hfq chaperone activity and sRNA stability (Zhang et al, 2013). The lateral surface of Hfq also interacts with
sRNAs and is important for sRNA stability as well as providing a favourable surface for nucleating pairing (Panja et al, 2013; Sauer et al, 2012; Sauer & Weichenrieder, 2011); however this surface of Hfq is dispensable for some systems. Unlike the proximal surface, the distal RNA-binding site is not an absolute requirement for sRNA-dependent regulation and in some cases may simply serve as a way to tether Hfq to target mRNAs (Sauer, 2013; Zhang et al, 2013). The role of each RNA-binding surface of Hfq in sRNA-independent regulation has not been studied.

We first measured the impact of Hfq binding face mutations on IS10 transposase expression and transposition. The finding that the proximal surface is only partially required for regulation suggested that Hfq is functioning independent of a trans-encoded sRNA. The moderate effect of the K56A mutation on Hfq is likely a result of reduced binding to site 2 in RNA-IN, which is supported by decreased regulation in the presence of the M2 mutations. Additionally, the lateral surface was not required for repressing transposase expression or transposition. Alone, the R17A phenotype does not exclude sRNA-dependent regulation. In the case of rpoS, multiple mutations to the lateral surface resulted in the strongest decrease in sRNA-mediated activation of rpoS expression (Panja et al, 2013). Additionally, the lateral surface was only important for about half of the sRNA/mRNA pairs tested previously (Zhang et al, 2013). However, a dispensable lateral surface is consistent with sRNA-independent regulation in the IS10 system. Unlike most sRNA-dependent regulation, the distal surface is critical for repressing transposase translation and mutations that block the interaction between the TIR of RNA-IN and the distal site on Hfq strongly de-repressed transposase expression in vivo. Additionally, an available distal surface on Hfq was required for blocking 30S ribosome binding to RNA-IN in vitro. We therefore suggest that sRNA-independent regulation by Hfq requires mRNA binding through the distal surface and not the proximal or lateral surfaces.

For Hfq to be an effective direct inhibitor of translation, we think some very specific requirements must be met. Firstly, there needs to be an Hfq binding site in the TIR of an mRNA. Interestingly, a recent survey of RNA sequences bound by Hfq in vivo included repeated trinucleotide motifs (ARN) that were frequently associated with the Shine-Dalgarno sequence. Notably, 18% of all Hfq associated mRNAs contained Hfq-
binding sites located within the TIR (Tree et al, 2014). Secondly, the Hfq binding site within the TIR should be a high affinity site to ensure in vivo binding. Thirdly, there must be sufficient available Hfq to act stochiometrically on TIRs. Given that Hfq is a highly expressed protein, this third factor might not appear to be a limitation. However, there is growing evidence that despite being highly abundant (5,000-10,000 hexamers per cell) (Ali Azam et al, 1999; Argaman et al, 2012; Kajitani et al, 1994) the amount of unbound Hfq at any given time might in fact be limiting for RNA binding (Azam & Ishihama, 1999; Hussein & Lim, 2011; Moon & Gottesman, 2011). Hfq has a large number of specific mRNA and sRNA targets and may also be sequestered through mostly non-specific DNA interactions (Updegrove, 2010 #313). Accordingly, for translational repression where a sustained interaction with an mRNA is required to block 30S ribosomal subunit binding, it is likely critical that Hfq bind with extremely high affinity to the TIR which might compensate for limited availability of Hfq. We found this to be the case in the IS10 system as Hfq bound the TIR with an affinity of approximately 0.2 nM. This represents one of the highest affinity interactions between Hfq and an mRNA (cf. ompA 1nM, ompC 0.9nM, ompF 4nM (Fender et al, 2010), sodB 0.3nM (Geissmann & Touati, 2004), rpoS ~50nM (Peng et al, 2014; Soper et al, 2011)). Moreover, we found that a moderate increase in expression of ChiX sRNA was sufficient to de-repress transposase expression. Given our evidence that: (i) ChiX possesses a distal Hfq binding site; (ii) its overexpression did not substantially influence RNA-IN steady-state levels, (iii) ChiX does not base-pair with RNA-IN, and (iv) ChiX overexpression reduces the amount of RNA-IN bound by Hfq in vivo, we think the results of the ChiX overexpression experiment are most easily explained by a ChiX-Hfq titrating mechanism. That is, despite the high affinity of site 1 in RNA-IN for Hfq, moderate overexpression of ChiX was sufficient to deplete the pool of available Hfq such that there were insufficient amounts to repress RNA-IN translation. The implications of Hfq titration by ChiX are discussed below.

3.3.2 Hfq regulation of IS10 in single and multi-copy

All of the in vivo experiments reported in this work were performed under conditions where the naturally occurring antisense RNA (RNA-OUT) was not produced. This was
meant to mimic a situation where IS10 is in single copy, where the antisense RNA has little to no effect on transposase expression (Kleckner, 1990). The HH104 mutation was introduced to the single-copy IS10 elements to increase transposase transcription to detectable levels, and although this mutation also eliminates the small amount of cis-encoded RNA-OUT there would be little antisense control to begin with. We propose that under these conditions Hfq acts in a stoichiometric manner to limit RNA-IN translation (Figure 3.10). However, when IS10 is present on a multi-copy plasmid stoichiometric action of Hfq on RNA-IN might not be sufficient to limit translation because of the increase in the number of RNA-IN transcripts. In support of this, when we compared Hfq regulation of a chromosomal IS10-lacZ fusion to a high-copy translational fusion, the extent of Hfq-mediated repression was attenuated from 13- to 3-fold (Figures 3.2A and 3.5). With a high-copy IS10 element, the MCI pathway comes into play and RNA-OUT becomes an important negative regulator of RNA-IN translation and stability (Case et al, 1990). We have previously shown that Hfq binds RNA-OUT and promotes its restructuring to expose sequences important for RNA-IN pairing (Ross et al, 2010). As Hfq facilitates RNA-IN:OUT pairing in a catalytic cycle, we think that when IS10 is in multi-copy, this mechanism would predominate over the ‘stoichiometric’ inhibition pathway. We eliminated RNA-OUT from the multi-copy IS10\textsubscript{1,242}-lacZ translational fusion so we could study antisense-independent regulation by Hfq, and our results show that stoichiometric repression can still occur albeit regulation is weaker for multi-copy IS10. We therefore think that Hfq is a negative regulator of IS10 regardless of copy-number. This model may also explain previous results where it was found that Hfq is a stronger negative regulator of transposition when IS10 is present in multi-copy compared to single-copy (Ross et al, 2010). Our model would predict that in this scenario Hfq is repressing predominantly by a catalytic mechanism involving Hfq-mediated RNA-IN:OUT pairing, which presumably is more efficient than the stoichiometric repression that is confined to the single copy situation.

This dual-model for Hfq repression may be applicable to other members of the Hfq regulon. In the absence of a cognate sRNA, we suggest that Hfq would repress translation of mRNAs containing an Hfq-binding site in the TIR by a stoichiometric
Figure 3.10. Model for dual role of Hfq in repressing IS10 transposase translation.

When IS10 is present in single copy (left side), transposase expression is not subject to antisense control by RNA-OUT. Hfq binding to the RBS in RNA-IN represses translation by preventing ribosome binding (‘stoichiometric’ repression). Multi-copy IS10 is subject to antisense control by increases concentrations of RNA-OUT (right panel). Hfq may still participate in stoichiometric repression by also facilitates antisense pairing in a ‘catalytic’ manner. Hfq binding to RNA-IN and RNA-OUT alters RNA secondary structure (not shown for RNA-IN), exposing sequences involved in pairing.
mechanism. The strength of this regulation would be governed primarily by the affinity of Hfq for that site relative to other cellular mRNAs. This model might also explain the conflicting data concerning regulation of *ompA* expression. Work in the Bläsi lab suggested that Hfq directly represses *ompA* translation by binding the TIR in a manner analogous to that described here for IS10 (Vytvytska et al, 2000). This is supported by several surveys of Hfq-binding mRNAs that have identified *ompA* as an Hfq-binding mRNA (Sittka et al, 2008; Tree et al, 2014; Zhang et al, 2003). The characterization of the stationary-phase sRNA MicA by the Wagner lab suggested that the primary role of Hfq in regulating *ompA* expression was to promote sRNA-mRNA pairing (Udekwu et al, 2005). Our model would combine both mechanisms and suggest that Hfq exerts some basal repression of *ompA* translation that is strengthened in stationary phase by promoting MicA pairing with *ompA*. Additionally, our model is applicable to the *sdhC* system where Hfq had an sRNA-independent effect on expression that is presumably a result of direct repression of translation (Desnoyers & Masse, 2012).

### 3.3.3 Hfq titration by the sRNA ChiX

Induction of Hfq-binding sRNAs can impinge on other Hfq-dependent post-transcriptional networks. In the simplest model, induction of an sRNA would provide enough sRNA molecules to bind all available Hfq and even compete with other Hfq-binding RNAs. Hfq titration was first proposed as a mechanism for OxyS repression of *rpoS* expression, which was later verified (Moon & Gottesman, 2011. ChiX overexpression also resulted in increased rpoS-lacZ expression, and this effect was proposed to be a result of Hfq titration (Mandin, 2010 #308; Zhang et al, 1998).

Additionally, overexpression of ArcZ in *Salmonella* was shown to have a pleiotropic effect on gene expression (altering expression of 757 genes) in part by decreasing the number of mRNAs bound to Hfq as well as specific competition with the sRNAs CyaR and InvR (Papenfort et al, 2009). Our studies with ChiX overexpression have provided another example of an sRNA sequestering sufficient amounts of Hfq to produce a biological effect; de-repression of RNA-IN translation.

ChiX (previously named SroB, RybC, and MicM) is a negative regulator of genes involved in chitobiose utilization. ChiX is constitutively expressed but its levels increase
substantially in stationary phase (Figueroa-Bossi et al, 2009; Papenfort & Vogel, 2014; Vogel et al, 2003). The fact that ChiX was the only sRNA of 14 screened to have an impact on expression of RNA-IN under anti-sense independent conditions fits fully with our model of stoichiometric inhibition resulting from Hfq binding the TIR of RNA-IN (site 1) through its distal binding site. ChiX is somewhat unique amongst *E. coli* sRNAs, as (according to our footprinting data) it possesses both distal and proximal Hfq binding sites. Moreover, in a recent study that compared the ability of a set of sRNAs to compete for Hfq binding, it was established that ChiX was at the top of the hierarchy, while SgrS was at the bottom. This same work also showed that the A-rich region of ChiX that we designated Hfq-I interacts with the distal surface of Hfq and provides important determinants for Hfq competition (Malecka et al, 2015). This fits fully with our data showing that ChiX overexpression titrated Hfq away from RNA-IN while SgrS did not.

ChiX is unique amongst sRNAs as it acts catalytically to repress its target mRNA (*chiP*, previously known as *ybfM*), and ChiX levels are regulated by an ‘anti-sRNA’, *chbBC* (Overgaard et al, 2009). It is therefore unexpected that ChiX levels would increase so dramatically during stationary phase. Given the current work, it is tempting to speculate that ChiX has a yet unidentified role during the transition to stationary phase; notably, ChiX constitutes 24-26% of Hfq-bound sRNAs during early stationary phase (Chao et al, 2012). Based on ChiX’s distinct interaction properties with Hfq, it is worth considering the possibility that ChiX expression could influence the entire Hfq regulon during the transition to stationary phase.

### 3.4 Materials and Methods

#### 3.4.1 Bacterial strains, phage, plasmids and oligonucleotides

All bacterial strains, phage, and plasmids used in this study are listed in Table S3.1 and oligonucleotides are listed in Table S3.2.

For mating out experiments, DBH33 was lysogenized with λDBH504 to create DBH331 (*hfq*+). P1 transduction was then performed to convert DBH331 to DBH337
(hfq). λDBH504 was created by crossing IS10<sup>HHi104</sup>-kan from pNK1223 onto λNK1039 in DBH60; kan<sup>R</sup> lysogens were selected by replica plating and then phage stocks (λDBH504) were prepared from these lysogens. For β-galactosidase assays with chromosomal IS10<sup>HHi104</sup>-lacZ, DBH107 was lysogenized with λRS271 (obtained from DBH90 via spontaneous phage release) to create DBH287. DBH287 was subjected to recombineering (details available upon request) to remove the G8 mutation creating DBH298. P1 transduction was then performed to convert DBH298 (hfq<sup>+</sup>) to DBH299 (hfq<sup>−</sup>).

The Hfq expression plasmids used for complementation experiments were made by amplifying the hfq gene (including the P3 promoter) from pDH700, pDH701, and pDH713 (Ross et al, 2013) with primers oDH518 and oDH519. The PCR product was digested with HaeIII and cloned into the XmnI/ScaI sites of pACYC184. The R17A mutation was first introduced into pDH700 by overlap PCR using primers oDH518, oDH519, oDH520, and oDH521. The PCR product was cloned into the XbaI/HindIII sites of pDH700 to make pDH874, and Hfq<sup>R17A</sup> was then subcloned into pACYC184 as above. Hfq was amplified from pDH904 with oDH184 and oDH479 to add a C-terminal 3xFLAG tag and this amplicon was cloned directly into XmnI/ScaI digested pACYC184.

The multi-copy IS10-lacZ translational fusion is a derivative of pNK2974 (Jain, 1995). First, IS10 was amplified with primers oDH502 and oDH503 and this amplicon was cloned into the EcoRI/HindIII sites of pNK2974 to produce pDH858 which contains the first 242nt of IS10R fused in frame to codon 10 of lacZ. All subsequent mutations were introduced into pDH858 using overlap PCR with primers oDH505 and oDH13 and the relevant mutagenic primers; R5 (oDH506, oDH507), M2 (oDH498, oDH499), and M5 (oDH508, oDH509). PCR products were digested with EcoRI and HindIII and cloned into the same sites in pDH858. M5 and M2 were originally identified as increased expression mutants in a transposase expression screen. The transposase gene used was derived from a library of sequences generated by mutagenic PCR.

Plasmids overexpressing sodB, ptsG, and maeA were constructed as previously described (Zhang et al, 2013). MC4100 genomic DNA served as a template for PCR with
the following primers: oDH558 and oDH559, sodB; oDH560 and oDH561, ptsG; oDH562 and oDH563, maeA. PCR products were digested with AatII and EcoRI and cloned into the same sites of pDH765.

### 3.4.2 Hfq footprinting and EMSA

*In vitro* transcription templates were generated by PCR using plasmids pDH866 (IN-160), pDH868 (IN_M2-160), pDH875 (IN_M5-160) and primers oDH515 (IN-160 and IN_M2-160) or oDH510 (IN_M5-160) with oDH199. RNA-IN was generated by *in vitro* transcription and internally labeled with [α-32P]-UTP (for EMSA) or 5'-labeled with [γ-32P]-ATP (for footprinting) as previously described (Ross et al., 2013). Wild-type Hfq was purified by heat treatment and poly(A) affinity purification, and his-tagged Hfq variants were purified by Ni<sup>2+</sup>-IMAC as previously described (Ross et al., 2013). RNA-footprinting was performed as previously described (Ross et al., 2013, except that reactions were in 1X RNA Structure Buffer (Ambion) and Pb<sup>2+</sup> footprinting used 10 mM Lead(II)Acetate (Sigma-Aldrich) for 3 min at ambient temperature which was stopped by addition of EDTA to a final concentration of 50 mM. Following ethanol precipitation, RNA footprinting samples were resuspended in denaturing load dye (95% [v/v] formamide, 0.5X TBE, 3% [w/v] xylene cyanol) and resolved on a 10% polyacrylamide gel containing 7M urea. EMSA was performed essentially as described (Ross et al., 2013) except that binding reactions included 20 ng/µL total yeast RNA (Ambion). Apparent dissociation constants were determined as previously described (Ross, 2013 #28).

The ChiX *in vitro* transcription template was generated with primers oDH528 and oDH529 with a genomic DNA template. Lead footprinting was performed as above. A 3.5 µL aliquot of ChiX, ChiX-Hfq, or ChiX-Hfq-RNA-IN (700 nM) was removed from binding reactions, mixed with native load dye (20 mM Tris-HCl, pH 7.5, 10 mM DTT, 100 mM KCl, 30% glycerol [v/v], 0.05% bromophenol blue [w/v]), and resolved on a 6% polyacrylamide TBE gel.
3.4.3 β-galactosidase assays

Cells were grown in LB supplemented (where necessary for plasmid selection) with ampicillin (100 µg/mL) and tetracycline (10 µg/mL). Saturated overnight cultures were used to seed subcultures (1:40 dilution), which were grown to mid-log phase (OD$_{600} = 0.4$–0.6). IPTG (1 mM) was added to the subculture to induce sRNA expression. The Miller assay was performed as previously described (Ross et al, 2010).

3.4.4 Conjugal mating out assay

The mating out assay was performed essentially as previously described (Ross et al, 2010). Plasmids encoding the Hfq variants (pDH904, pDH905, pDH906, and pDH907) or a vector control (pDH900) were transformed into the donor strain DBH337 (hfq$^-$; contains IS10$^{HH104}$-kan lysogen) and plated on M9-glucose supplemented with thiamine (1 µg/mL), arginine (40 µg/mL), kanamycin (50 µg/mL) and tetracycline (15 µg/mL). Donor colonies were grown overnight to saturation in LB supplemented with kanamycin (50 µg/mL) and tetracycline (15 µg/mL) and the recipient strain (HB101) was grown in LB supplemented with streptomycin (150 µg/mL). Donor and recipient strains were subcultured in LB without antibiotics, and then grown and mixed for mating as previously described (Ross et al, 2010). Mating was stopped by vigorous vortexing after 1 hr and 1 mL of mating mixture was washed and then serially diluted in saline (0.85% [w/v] NaCl). Cells were plated on M9-glucose supplemented with thiamine, leucine (40 µg/mL) and streptomycin (150 µg/mL) or streptomycin plus kanamycin (50 µg/mL) for ‘total exconjugates’ and ‘hops’ respectively. The transposition frequency was calculated by dividing the number of Sm$^R$Kan$^R$ colonies by Sm$^R$ colonies (‘hops’ per ‘exconjugate’).

3.4.5 RNA extraction and primer extension analysis

Cells were grown in LB supplemented with ampicillin (100 µg/mL) to OD$_{600} =$ 0.6 at which time 600 µL of cells were added to 300 µL of RNA lysis buffer (1.5% [w/v] SDS, 300 mM sodium acetate, 30 mM EDTA) and boiled for 1 min. Samples were
chilled on ice for 30s and then sequentially extracted twice with acid phenol (pH 4.3), once with phenol:chloroform:isoamyl alcohol (25:24:1) and once with 2-butanol followed by ethanol precipitation. Residual genomic DNA was removed with TURBO DNase (Ambion) prior to primer extension analysis. 5 or 10 µg of total RNA (Figure 5B, 8C and 9 respectively) was subject to primer extension analysis with 5\(^{32}\)P-labeled oDH511 (RNA-IN) and oDH482 (lpp) and SuperScript III (Invitrogen) according to the manufacturer’s instructions.

**3.4.6 Toeprinting**

The 30S ribosomal subunit was prepared as previously described (Fechter et al, 2009). *In vitro* transcribed RNA (2 pmol) was annealed to 5\(^{32}\)P-labeled oDH511 (RNA-IN), oDH482 (lpp) and oDH555 (usg) in buffer SB (10 mM Tris-acetate, pH 7.6, 1 mM DTT, 100 mM potassium acetate) by heating to 95°C for 1 min followed by snap-cooling on ice for 2 min. While on ice, magnesium acetate was added to a final concentration of 10 mM and dNTPs to 0.5 mM. Reactions were then incubated at 37°C for 5 min. Hfq (0.5-4 pmol of hexamers) or buffer was added to reactions which were incubated for another 15 min at 37°C, followed by addition of 30S ribosome (3.6 pmol for RNA-IN and usg, 2.7 pmol for lpp) and incubation at 37°C for 5 min. Initiator fMet-tRNA (10 pmol; Sigma-Aldrich) was added and reactions were incubated for a further 15 min before addition of 200U of SuperScript II (Invitrogen) and a final incubation of 10 min at 37°C. Reactions were stopped by addition of 100 µL of stop solution (50 mM Tris-HCl, pH 7.5, 0.1% SDS [w/v], 10 mM EDTA) followed by phenol:chloroform:isoamyl alcohol extraction and ethanol precipitation. Samples were resuspended in denaturing load dye (95% [v/v] formamide, 0.5X TBE, 3% [w/v] xylene cyanol) and resolved on a 10% polyacrylamide gel containing 7M urea. Dried gels were exposed to a phosphorimager storage screen, imaged with a Storm imager and quantitated with ImageQuant (GE Healthcare).
3.4.7 Hfq-RNA immunoprecipitation (RIP)

DBH337 (hfq^-; contains IS10^HH104^-kan lysogen) was transformed with plasmids expressing untagged Hfq (pDH904) or Hfq with a C-terminal 3xFLAG tag (pDH909). Cells were grown to mid-exponential phase (OD_600 = 0.5) in LB supplemented with tetracycline (15 µg/mL) at which point 50 OD_600 of cells was collected by centrifugation and washed once in TBS (50 mM Tris-HCl, pH 7.4, 150 mM NaCl) and resuspended in 400 µL of lysis buffer (20 mM Tris-HCl, pH 8.0, 150 mM KCl, 1 mM MgCl2) with 40U of Rnasin (Promega) and 2U of TURBO DNase (Ambion). Cells were mixed with 400 µL of zirconia/silica beads (0.1 mm, BioSpec) and lysed by vortexing (30 s burst, 30 s ice; 10 cycles) after which 800 µL of lysis buffer was added followed by centrifugation (10 min, 13,500 RCF, 4°C). An aliquot of the cleared lysate (100 µL) was phenol extracted and ethanol precipitated (‘Input RNA’). ANTI-FLAG® M2 magnetic beads (25 µL packed resin; Sigma-Aldrich) were added to 800 µL of cleared lysate and samples were incubated at 4°C with rotation for 4 h. Beads were washed 5 times with 1 mL of lysis buffer, and resuspended in 400 µL of lysis buffer. Hfq-bound RNA (‘IP RNA’) was recovered by phenol extraction and ethanol precipitation. Following precipitation, residual DNA was removed with TURBO DNase (Ambion) and samples were ethanol precipitated and finally resuspended in DEPC treated ddH_2O. RNA concentration was determined with a NanoPhotometer (Implen).

For the RIP with WT and mutant IS10-lacZ (Figure 3.7B) DBH12 was transformed with pDH909 (Hfq-3xFLAG) and pDH866 (IN^WT-lacZ) or pDH875 (IN^M5-lacZ). Cells were grown in LB supplemented with tetracycline and ampicillin (100 µg/mL) and IP was performed as above.

The RIP with ChiX overexpression used DBH337 (hfq^-; contains IS10^HH104^-kan lysogen) transformed with pDH909 (Hfq-3xFLAG) and pDH765 (ChiX) or pDH763 (vector). IP was performed as above. Total input RNA (10 µg) or RNA recovered from the IP (0.3 µg) was analyzed directly by northern blot (ChiX and 5S rRNA) or primer extension (RNA-IN and lpp).
3.4.8 RT-PCR

An RNA adapter (oDH486; 100 pmol) was ligated to input (10 µg) or IP (1 µg) RNA using T4 RNA ligase. Adapter-ligated RNA was purified with an RNeasy Mini Kit (Qiagen), eluted in 30 µL of DEPC ddH₂O, and 10 µL of RNA was converted to cDNA using an adapter specific primer (oDH352) and SuperScript III (Invitrogen) according to the manufacturers instructions. Enzyme was omitted for the no reverse transcription controls (-RT). Reverse transcription reactions were purified with a PCR purification kit (Qiagen) and eluted in 50 µL of ddH₂O. The first 160 nt of RNA-IN was detected by 28 cycles of PCR using 4 µL of cDNA as a template and primers oDH199 and oDH483 (WT) or oDH517 (M5). A portion of the 16S rRNA (nt 1071-1425) was detected by 18 cycles of PCR using primers oDH204 and oDH205. PCR reactions were analyzed on a 2% agarose TBE gel.

3.4.9 Northern Blot

3.5 µg of total RNA (Figure 3.8B) was denatured in load dye (95% [v/v] formamide, 0.5X TBE, 3% [w/v] xylene cyanol) and then separated on a 10% polyacrylamide gel containing 7M urea. RNA was electroblotted to a Hybond-N nylon membrane (Amersham) in 0.5X TBE at 200 mA for 1 hr. RNA was UV cross-linked to the membrane and then probed for SgrS with 5'³²P-labeled oDH298 in ULTRAhyb-oligo buffer (Ambion) at 42°C. ChiX and the 5S rRNA were detected by probing the membrane with an internally ³²P-labeled antisense RNA (templates were generated by PCR with genomic DNA template and primers oDH234 and oDH235 (5S rRNA) or oDH308 and oDH309 (ChiX)) in ULTRAhyb buffer (Ambion) at 68°C. Membranes were washed according to the manufacturer’s instructions. Northern blots were exposed to a phosphorimager storage screen and imaged with a STORM imager (GE Healthcare).

3.5 Supplemental Material
Table S 3.1. Strains, plasmids, and bacteriophage for Chapter 3

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBH12</td>
<td>MC4100 *hfq-1::ΩCm; Str&lt;sup&gt;R&lt;/sup&gt;Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>G. Storz; β-galactosidase assays and source of <em>hfq-1</em> allele for P1 transductions</td>
</tr>
<tr>
<td>DBH13</td>
<td>HB101 [F' <em>leu pro</em> ]; Str&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Mating out recipient</td>
</tr>
<tr>
<td>DBH33</td>
<td>NK5830 [<em>recA arg ΔlacproXIII nal&lt;sup&gt;R&lt;/sup&gt; rif&lt;sup&gt;R&lt;/sup&gt;/ F' lacpro</em>]</td>
<td>Parent strain for mating out donors</td>
</tr>
<tr>
<td>DBH60</td>
<td>C600</td>
<td>Used for plasmid-λ crosses</td>
</tr>
<tr>
<td>DBH90</td>
<td>DBH33 λRS271 (G8, HH104); Kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Chromosomal IS10-lacZ translation fusion marked with Kan&lt;sup&gt;R&lt;/sup&gt;; HH104 mutation increases transposase transcription and G8 mutation increases RNA-OUT expression to maintain antisense control</td>
</tr>
<tr>
<td>DBH107</td>
<td>MC4100; Str&lt;sup&gt;R&lt;/sup&gt;</td>
<td>β-galactosidase assays</td>
</tr>
<tr>
<td>DBH287</td>
<td>DBH107 λRS271 (G8; HH104); Str&lt;sup&gt;R&lt;/sup&gt;Kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Parent strain</td>
</tr>
<tr>
<td>DBH298</td>
<td>DBH107 λRS271 (HH104); Str&lt;sup&gt;R&lt;/sup&gt;Kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Chromosomal IS10-lacZ translational fusion marked with Kan&lt;sup&gt;R&lt;/sup&gt;, G8 removed by recombineering, no RNA-OUT; β-galactosidase assays (wt)</td>
</tr>
<tr>
<td>DBH299</td>
<td>DBH298 *hfq-1::ΩCm; Kan&lt;sup&gt;R&lt;/sup&gt;Str&lt;sup&gt;R&lt;/sup&gt;Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>β-galactosidase assays (hfq&lt;sup&gt;−&lt;/sup&gt;)</td>
</tr>
<tr>
<td>DBH313</td>
<td>MRE600 [F' Δrna]</td>
<td>Steitz Lab; Strain for 30S ribosome purification</td>
</tr>
<tr>
<td>DBH331</td>
<td>DBH33 λDBH504 (HH104); Kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Chromosomal IS10-Kan HH104; mating out donor strain, no RNA-OUT (wt)</td>
</tr>
<tr>
<td>DBH337</td>
<td>DBH331 *hfq-1::ΩCm; Kan&lt;sup&gt;R&lt;/sup&gt;Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Mating out donor strain and source of native RNA-IN for RIP experiments, no RNA-OUT (hfq&lt;sup&gt;−&lt;/sup&gt;) Cloning and plasmid propagation</td>
</tr>
<tr>
<td><strong>DH5α</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pDH858</td>
<td>pUC119 derived; IS10&lt;sub&gt;1-242&lt;/sub&gt;-lacZ translational fusion; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>WT IS10-lacZ translational fusion, expresses RNA-OUT</td>
</tr>
<tr>
<td>pDH866</td>
<td>pDH858 + R5</td>
<td>WT RNA-IN, RNA-OUT destabilized by R5 mutation</td>
</tr>
<tr>
<td>pDH868</td>
<td>pDH866 + M2</td>
<td>Mutated Hfq binding site 2 on RNA-IN</td>
</tr>
</tbody>
</table>
pDH875  pDH866 + M5  Mutated Hfq binding site 1 on RNA-In
pDH700  pWKS30-p3-Hfq\textsuperscript{WT}; Ap\textsuperscript{R}  Parent plasmid; expression vector for Hfq\textsuperscript{WT}
pDH701  pDH700-Hfq\textsuperscript{K56A}  Parent plasmid; expression vector for Hfq\textsuperscript{K56A}
pDH713  pDH700-Hfq\textsuperscript{Y25A}  Parent plasmid; expression vector for Hfq\textsuperscript{Y25A}
pDH874  pDH700-Hfq\textsuperscript{R17A}  Parent plasmid; expression vector for Hfq\textsuperscript{R17A}
pDH900  pACYC184 with ScaI/XmnI fragment removed; Tet\textsuperscript{R}Cm\textsuperscript{S}  Vector control for Hfq
pDH904  pDH900-p3-Hfq\textsuperscript{WT}  Expression vector for Hfq\textsuperscript{WT}
pDH905  pDH900-p3-Hfq\textsuperscript{Y25A}  Expression vector for Hfq\textsuperscript{Y25A}
pDH906  pDH900-p3-Hfq\textsuperscript{K56A}  Expression vector for Hfq\textsuperscript{K56A}
pDH907  pDH900-p3-Hfq\textsuperscript{R17A}  Expression vector for Hfq\textsuperscript{R17A}
pDH909  pDH900-p3-Hfq\textsuperscript{WT} -3xFLAG  Expression vector for Hfq\textsuperscript{WT} - 3xFLAG
pDH763  pBR-pLlacO; Ap\textsuperscript{R}  Vector control for sRNA expression
pDH764  pBR-pLlacO-sgrS; Ap\textsuperscript{R}Tet\textsuperscript{R}  SgrS expression
pDH765  pBR-pLlacO-chiX; Ap\textsuperscript{R}Tet\textsuperscript{R}  ChiX expression
pDH766  pBR-pLlacO-RybB; Ap\textsuperscript{R}Tet\textsuperscript{R}  RybB expression
pDH767  pBR-pLlacO-FnrS; Ap\textsuperscript{R}Tet\textsuperscript{R}  FnrS expression
pDH768  pBR-pLlacO-MicC; Ap\textsuperscript{R}Tet\textsuperscript{R}  MicC expression
pDH769  pBR-pLlacO-RydC; Ap\textsuperscript{R}Tet\textsuperscript{R}  RydC expression
pDH770  pBR-pLlacO-MgrR; Ap\textsuperscript{R}Tet\textsuperscript{R}  MgrR expression
pDH771  pBR-pLlacO-RprA; Ap\textsuperscript{R}Tet\textsuperscript{R}  RprA expression
pDH772  pBR-pLlacO-RyeB; Ap\textsuperscript{R}Tet\textsuperscript{R}  RyeB expression
pDH773  pBR-pLlacO-CyaR; Ap\textsuperscript{R}Tet\textsuperscript{R}  CyaR expression
pDH774  pBR-pLlacO-MicF; Ap\textsuperscript{R}Tet\textsuperscript{R}  MicF expression
pDH775  pBR-pLlacO-GlmY; Ap\textsuperscript{R}Tet\textsuperscript{R}  GlmY expression
pDH776  pBR-pLlacO-MicA; Ap\textsuperscript{R}Tet\textsuperscript{R}  MicA expression
pDH777  pBR-pLlacO-GcvB; Ap\textsuperscript{R}Tet\textsuperscript{R}  GcvB expression
pDH920  pBR-pLlacO-sodB\textsubscript{56 to +244}; Ap\textsuperscript{R}Tet\textsuperscript{R}  sodB expression
pDH921  pBR-pLlacO-ptsG\textsubscript{102 to +230}; Ap\textsuperscript{R}Tet\textsuperscript{R}  ptsG expression
pDH922  pBR-pLlacO-maeA\textsubscript{73 to +241}; Ap\textsuperscript{R}Tet\textsuperscript{R}  maeA expression

\textbf{\lambda phage}
\begin{itemize}
  \item \lambda RS271  Source of IS10-Kan HH104 G8
  \item \lambda NK1039  Encodes his operon
  \item \lambda DBH504  \lambda NK1039 with IS10-Kan HH104
\end{itemize}
\textbf{plasmid-\lambda crosses}
\begin{itemize}
  \item \lambda NK1039  Product of recombination between \lambda NK1039 and pNK1223
\end{itemize}
<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5' → 3')</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>oDH502</td>
<td>GGATCGGAATTCCGATCTTCCCTGATGATCCCCCTAATGTGATTTTG</td>
<td>Fwd primer for making pDH858</td>
</tr>
<tr>
<td>oDH503</td>
<td>CCCCCCAAGCGTTGCGCGGAAGTTCCGTTGAGAGTGGTGGTGGTAAAC</td>
<td>Rev primer for making pDH858</td>
</tr>
<tr>
<td>oDH505</td>
<td>GGATCGGAATTCCGATCTTCC</td>
<td>IS10R Fwd primer used for introducing mutations into pDH866</td>
</tr>
<tr>
<td>oDH13</td>
<td>CCAGTCAGACGTTGTAACAC</td>
<td>LacZ reverse primer used for introducing mutations into pDH866</td>
</tr>
<tr>
<td>oDH506</td>
<td>CCAGTCAGACGTTGTAACAC</td>
<td>Fwd primer for making IN-160 an IN&lt;sup&gt;M2&lt;/sup&gt;-160 ITR templates; contains T7 core promoter</td>
</tr>
<tr>
<td>oDH507</td>
<td>CCAGTCAGACGTTGTAACAC</td>
<td>Rev primer for making IN-160 ITR templates; Rev primer for PCR amplification of RNA-IN cDNA</td>
</tr>
<tr>
<td>oDH508</td>
<td>CCAGTCAGACGTTGTAACAC</td>
<td>Fwd primer for making IN&lt;sup&gt;M5&lt;/sup&gt;-160 ITR template; contains T7 core promoter</td>
</tr>
<tr>
<td>oDH509</td>
<td>CCAGTCAGACGTTGTAACAC</td>
<td>Primer for making RNA-IN-100 cDNA (primer extension and toeprinting)</td>
</tr>
<tr>
<td>oDH510</td>
<td>CCAGTCAGACGTTGTAACAC</td>
<td>Fwd primer for making lpp ITR template; contains T7 promoter sequence</td>
</tr>
<tr>
<td>oDH511</td>
<td>CCAGTCAGACGTTGTAACAC</td>
<td>Reverse primer for making lpp ITR template; also used for lpp primer extension and toeprinting</td>
</tr>
<tr>
<td>oDH512</td>
<td>CCAGTCAGACGTTGTAACAC</td>
<td>Fwd primer for Hfq gene</td>
</tr>
<tr>
<td>oDH513</td>
<td>CCAGTCAGACGTTGTAACAC</td>
<td>Rev primer for adding C-terminal 3xFLAG tag</td>
</tr>
<tr>
<td>oDH514</td>
<td>CCAGTCAGACGTTGTAACAC</td>
<td>RNA adapter; ligated to total RNA for reverse transcription</td>
</tr>
<tr>
<td>oDH515</td>
<td>CCAGTCAGACGTTGTAACAC</td>
<td>Fwd primer for RT-PCR of RNA-</td>
</tr>
</tbody>
</table>
oDH517  GCGAAAAGTCAGTCTGAGAC  IN  Fwd primer for RT-PCR of RNA-IN
oDH204  CGTGGTTGTGAATGGTTGGTTAAGT  Primer pair for RT-PCR of 16S
oDH205  AACCCACTCCCCATGGTGTGACGGGC  rRNA (nt 1071-1425)
oDH518  GCCAGCTGCGCAAGG  Primer pair for amplifying MCS of
oDH519  TAATGCAGCTGCGACGAG  pWKS30 derived plasmids; used
for introducing R17A mutation into
pDH700

oDH520  GCACCTCGTGCGGAAGCTGTTCC  Mutagenic primers for introducing
oDH521  GGAACACGTTCCGACGAGTGC  R17A mutation into pDH700
oDH352  TTTCATCAAAGCCAGCAAGGC  RNA adapter specific primer used
for cDNA synthesis
oDH353  CAAAACCGAAGTGCTTCTCATACG  RNA adapter specific nested
primer used for PCR of total RNA-IN cDNA

oDH234  TAATACGACTCTAGATAGCCTGGCAGT  Fwd primer for making antisense
TCCCTACTCTCG  5S rRNA (rrfA) ITR template for
Northern blot; contains T7
promoter sequence
oDH235  CGGCAGTAGCCTCAGAGGGAG  Reverse primer for above
oDH308  TAATACGACTCTAGATAGCCTGGCAGT  Fwd primer for making antisense
TGGCCAATATCGCTATTG  ChiX ITR template for Northern
blot; contains T7 promoter
sequence

oDH309  ACAACCGTCGCTTAAAGTGACG  Reverse primer for above
oDH298  CGGGTAAGTTCACGACATAC  Oligonucleotide for SgrS Northern
blot
oDH528  TAATACGACTCTAGACACCGTCTTAA  Fwd primer for making ChiX ITR
AGTGAC  template; contains T7 promoter
sequence
oDH529  AAAAAATGGCCAATATCGCTATTGCG  Reverse primer for above
oDH554  TAATACGACTCTAGATAGCCTGGCAGT  Fwd primer for making usg ITR
AACATGGCG  template; contains T7 promoter
sequence
oDH555  CCCACAGCGCCAGTTGC  Reverse primer for above; also
used for toeprinting
oDH558  AATACTGACGTCATACGCACAATAAGG  Primer pair for amplifying sodB for
CTATTG  overexpression in pDH920
oDH559  AATACTGACGTCATACGCACAATAAGG  Primer pair for amplifying ptsG for
AAAG  overexpression in pDH921
oDH560  AATACTGACGTCATACGCACAATAAGG  Primer pair for amplifying maeA
TTAG
oDH563 AATACGTGAATTCGGATGTTACGCAGGTA for overexpression in pDH922
GATG
Prior to measuring β-galactosidase activity of IS10-lacZ (Figure 3.2a), an equivalent number of cells (0.4 OD<sub>600</sub>) were pelleted and frozen at -80°C. Pellets were resuspended in SDS load-mix (2% SDS, 10% glycerol, 50 mM Tris-HCl pH 6.8, 0.25% bromophenol blue and 0.8 M β-mercaptoethanol) and boiled for 5 minutes. Samples were separated on a 14% SDS-PAGE gel and then transferred to a PVDF membrane (Roche). Hfq was detected by Western Blot using a polyclonal rabbit anti-Hfq antibody and a Pierce ECL 2 Western Blotting kit (Thermo Scientific). Purified Hfq (0.6 pmol) was loaded as a positive control.
Figure S 3.2. Mutations to the ribosome-binding site (RBS) of RNA-IN prevent Hfq binding.

$^{32}$P-labeled RNA-IN-160 (150 nM) was incubated -/+ 400 nM Hfq (hexamers) before treatment with RNase V1. Untreated controls and a RNase T1 sequencing ladder (lane 1) are shown. Nucleotide numbering is relative to the translational start. Four positions that are strongly protected from cleavage on wild-type RNA-IN (lanes 4-5) and IN$^{M2}$ (lanes 12-13) are indicated with red boxes. The corresponding positions that are not protected on IN$^{M5}$ are indicated with dashed red boxes. Note that the *in vitro* transcript for IN$^{M5}$ contains an extra 5’nt, resulting in a 1nt difference in the migration of RNA fragments (indicated by black arrows). Note that certain lanes from this gel were removed for clarity (indicated by black vertical lines).
Figure S 3.3. Toeprint analysis of 30S ribosome binding to \textit{lpp} or \textit{usg} mRNA in the presence of Hfq.

(A) Unlabeled \textit{lpp} (-38 to +144 relative to the translational start) or (B) \textit{usg} (-50 to +50 relative to the translational start) mRNA was incubated with the indicated concentrations of purified Hfq before addition of the 30S ribosomal subunit (\textit{lpp}, 270 nM; \textit{usg}, 360 nM) and fMet-tRNA (5 \textmu{}M). CAG refers to sequencing lanes produced with the same RNA used for toeprinting. The toeprint signal at positions +15/+16 relative to the start codon is indicated as well as a band used to normalize loading for quantitation of the \textit{lpp} toeprint (+82).
Figure S3.4. Specificity of Hfq:RNA-IN IP and size distribution of RNA-IN molecules containing a 5’end.

(A) Hfq western blot of fractions from Hfq<sup>WT</sup> and Hfq<sub>WT</sub>-3xFLAG immunoprecipitations. Aliquots of total cellular lysate (lanes 2 and 3), supernatant from IP reactions (lanes 3 and 6), or the ANTI-FLAG® M2 Magnetic Beads following washes (lanes 4 and 7) were subject to western blot as described in Figure S3.1. Purified Hfq (0.15 pmol) was loaded as a positive control (lane 1). The tagged and untagged Hfq monomers are indicated. (B) cDNA from total RNA (lanes 1 and 3) or IP reactions (lanes 2 and 5) was amplified by 24 cycles of PCR using an RNA-IN specific forward primer and a nested RNA Adapter specific reverse primer. PCR products were then resolved on a 2% Agarose/TBE gel and stained with EtBr. A DNA ladder (lane 5) was used to determine the size distribution of PCR products which is indicated with dashed horizontal lines. (C) Schematic of RT-PCR used for Hfq IP. Total RNA from the RIP was ligated to an RNA adapter (black box). An adapter specific primer was for cDNA synthesis, yielding a mixed population of RNA species. The size distribution of RNA-IN molecules with an intact 5’end (B) was determined by amplifying PCR with an RNA-IN specific forward primer (blue arrow) that anneals to the extreme 5’end and a nested adapter specific primer (black arrow). PCR shown in Figure 7 used an RNA-IN specific forward primer (blue) and an RNA-IN specific reverse primer (dashed black arrow) to amplify the first 160nt of RNA-IN.
Accordingly, the RNA-IN signal shown in Figure 3.7 originates from RNA-IN molecules containing at least the first 160nt of RNA-IN.
150

UT

- WT Y25A K56A

- WT Y25A K56A

Pb^{2+}

82

Hfq-III

78 (Proximal)

\( \rho \)-independent terminator

52

Hfq-II

48 (Proximal)

33

Hfq-I

(Distal)

24

1 2 3 4 5 6 7 8
Figure S 3.5. Hfq determinants for ChiX binding.

$5'\text{P}$-labeled ChiX (140 nM) was incubated with WT and mutant variants of his-tagged Hfq (380 nM) before limited cleavage with Pb$_2^+$ (lanes 5-8). Untreated controls are shown (lanes 1-4). The location of the two Hfq-binding sites identified in Figure 3.9C (Hfq-I and Hfq-II) are highlighted along with a third site (Hfq-III) that corresponds to the polyU tail after the $\rho$-independent transcriptional terminator. Sites are defined as interacting with the distal surface if they are protected by Hfq$^{WT}$ and Hfq$^{K56A}$ but not Hfq$^{Y25A}$. Likewise, sites that show reduced protection with Hfq$^{K56A}$ compared to Hfq$^{WT}/Hfq^{Y25A}$ are defined as interacting with the proximal surface of Hfq.
Figure S 3.6. Impact of mRNA overexpression on transposase expression.

*hfq*− cells containing the chromosomal IS101-337-\(lacZ\) translational fusion (DBH298) were transformed with a plasmid expressing ChiX (pDH765), sodB-56 to +244 (pDH920), ptsG-102 to +230 (pDH921), maeA-73 to 241 (pDH922) or a vector control (pDH763). Transformants were grown to mid-exponential phase in LB media and β-galactosidase activity was measured by the Miller assay. Error bars show the standard error on the mean for two independent experiments (n=6) and the relative expression is shown above the graph, where transposase-\(lacZ\) expression in the presence of vector was set to 1.
3.6 References


Jain C (1995) IS10 antisense control in vivo is affected by mutations throughout the region of complementarity between the interacting RNAs. *Journal of molecular biology* **246**: 585-594


Kleckner N (1990) Regulating tn10 and is10 transposition. *Genetics* **124**: 449-454


Chapter 4

4 A cis-encoded sRNA, Hfq and mRNA secondary structure act independently to suppress IS200 transposition

4.1 Introduction

Small non-coding RNAs (sRNAs) play an important role in regulating many physiological processes in bacteria, including but not limited to metabolism, stress response, and virulence (reviewed in (Papenfort & Vogel, 2010; Papenfort & Vogel, 2014; Storz et al, 2011; Waters & Storz, 2009)). Most sRNAs regulate gene expression through complementary base-pairing with target mRNAs, which usually affects translation and often transcript stability. The best studied class of sRNAs are expressed in trans relative to their target mRNA, and accordingly have only partial sequence complementarity. The chaperone protein Hfq is important for the function of most trans-sRNAs, protecting the sRNA from degradation and facilitating pairing between sRNAs and their target(s). Conversely, cis-encoded sRNAs (also called antisense RNA, asRNA) are expressed from the same loci as mRNAs on the opposite strand of DNA. This results in perfect and usually extended complementarity between asRNAs and their target mRNAs. Hfq is typically thought to be dispensable for asRNA regulation (Georg & Hess, 2011; Thomason & Storz, 2010), although there are a few systems where this is not the case (Opdyke et al, 2004; Ross et al, 2013).

The first sRNAs discovered in bacteria were asRNAs involved in plasmid and transposon copy-number control (Tomizawa & Itoh, 1981) (Simons et al, 1983). In the case of IS10, translation of the transposase mRNA (RNA-IN) is inhibited by the cis-encoded asRNA, RNA-OUT. Pairing between these RNAs initiates between the 5’end of RNA-IN and the terminal loop domain of RNA-OUT. Propagation of the paired species

---

to ultimately include 35 intermolecular base-pairs blocks 30S ribosome binding to RNA-IN (Kittle et al, 1989; Ma & Simons, 1990). Since RNA-OUT can act in trans on all copies of IS10 in a cell, the strength of antisense regulation increases with IS10 copy-number and accordingly plays an important role in limiting transposition (Simons & Kleckner, 1983). Identification of new functional sRNAs has been aided by the development of RNA-Seq coupled with Hfq immunoprecipitation (Hfq-IP) (Sharma & Vogel, 2009). By sequencing sRNAs that interact with Hfq, it is possible to separate putative functional sRNAs from spurious transcription products. One surprising observation from these Hfq-IP experiments is that Hfq interacts with a number of cis-encoded sRNAs. The first study to use Hfq-IP for identifying sRNAs in Salmonella enterica serovar Typhimurium (hereafter Salmonella) found that about 3% of Hfq-bound RNA mapped antisense to protein coding regions (Sittka et al, 2008). More strikingly, asRNAs made up the second largest class (25%) of Hfq-binding sRNAs in Mycobacterium smegmatis although the significance of this is unclear as no Hfq orthologs have been identified in Mycobacterium species so far (Li et al, 2013). Escherichia coli may express up to 300 functional asRNAs, although only 67 were detected in an Hfq-IP (Bilusic et al, 2014; Lybecker et al, 2014). Hfq may therefore play a previously unappreciated role in antisense regulation. Alternatively, the subset of asRNAs that interact with Hfq may be trans-sRNAs that just happen to be expressed antisense to protein coding genes.

We have previously shown that Hfq facilitates antisense pairing between the IS10 transposase mRNA (RNA-IN) and the cis-encoded sRNA, RNA-OUT (Ross et al, 2013). We were interested in determining if Hfq regulated other transposons by a similar mechanism and accordingly searched Hfq-IP data sets for evidence of Hfq-interacting RNAs that are antisense to transposase mRNAs. IS200 encodes an sRNA (STnc490) that is antisense to 90 nucleotides (nt) of the transposase 5’ untranslated region (5’UTR) in Salmonella (Kroger et al, 2012; Sittka et al, 2008). The closely related IS1541 element from Yersinia pestis also expresses STnc490 (Yan et al, 2013). Promoters for asRNAs can arise stochastically, but the conserved expression of STnc490 suggests this is a functional asRNA (Raghavan et al, 2012).
IS200 elements are ubiquitous in Enterobacteriaceae and have been identified throughout Eubacteria and Archaea (Beuzon & Casadesus, 1997; Beuzon et al, 2004; Filee et al, 2007; Gibert et al, 1990). IS200 was first identified as a polar insertion mutant in the *hisD* gene of *Salmonella* (*hisD984*; (Lam & Roth, 1983)). Repeated attempts to measure IS200 transposition under various laboratory conditions were unsuccessful, and environmental samples of *Salmonella* collected 30 years apart showed no evidence of transposition (Beuzon et al, 2004; Casadesus & Roth, 1989; Schiaffino et al, 1996). However, IS200 does transpose during long-term stab culture and there is evidence that the closely related IS1541 element is active during mouse infection by *Y. pestis* (Cornelius et al, 2009; Lam & Roth, 1983). Taken together these observations have led to the conclusion that IS200 is a mostly dormant transposable element (Beuzon & Casadesus, 1997; Bisercic & Ochman, 1993). A reasonable presumption would be that most IS200 elements are inactive remnants of the active transposon. However, sequence comparison of ‘genomic’ IS200 elements and rare transposition products revealed that the sequence of ‘active’ and ‘inactive’ elements is almost identical (Beuzon & Casadesus, 1997). It therefore seems likely that the ‘native’ state of IS200 elements is ‘off’ for transposition although specific conditions might lead to sporadic transposition.

Transposition of many bacterial transposons is limited by the expression of the transposase genes they encode (Kleckner, 1990; Morisato et al, 1983; Nagy & Chandler, 2004). IS200, the smallest fully autonomous insertion sequence known, contains a single open reading frame (ORF) that encodes a transposase protein, TnpA (Figure 4.1A). Transcription of the *tnpA* gene is limited by an intrinsically weak promoter and a bi-directional rho-independent terminator in the ‘left end’ that protects against impinging transcription (Beuzon & Casadesus, 1997; Beuzon et al, 1999). The left end contains a second inverted repeat that comprises a portion of the 5’UTR of *tnpA*. This is most likely a cis-regulatory element that represses *tnpA* translation by sequestering the Shine-Dalgarno (SD) sequence (Beuzon et al, 1999). Antisense control of *tnpA* expression would therefore be an additional level of regulation for transposase protein.

In the current work we asked if IS200 transposase expression is down-regulated by the cis-encoded sRNA in *E. coli*. We show that transposase expression is strongly
Figure 4.1. Schematic of IS200.

(A) IS200 is 707 base-pairs in length. It contains a single protein coding gene (transposase; *tnpA*), transcription of which originates at about nt 40 (Beuzon & Casadesus, 1997); *tnpA* promoter elements have not been defined. The left end contains two internal inverted repeats (opposing arrows), one of which acts as a transcription terminator (nts 12-34) and the other (nts 69-138) was predicted to encode a stem-loop structure in the 5' UTR of the *tnpA* mRNA that sequesters the Shine-Dalgarno sequence. IS200 in *Salmonella* also expresses a 90 nt sRNA (art200, previously STnc490), which is perfectly complementary to the 5' UTR and the first three codons of *tnpA*. The transcription start site and 3' end for art200 in *Salmonella* (derived from RNA-Seq experiments) are shown but promoter elements were not previously defined (Kroger et al, 2012). (B) The DNA sequence of the first 200 nucleotides of IS200 is shown. The *tnpA* and art200 transcripts are shown in grey and black, respectively. Putative promoter elements for art200 are boxed and the *Salmonella* transcription start site (+1) is shown. The former were predicted using a position weight matrix (showing nucleotide identity of -10 and -35 promoter elements for *E. coli*) and the optimal spacing between -10 and -35.
elements in *E. coli* (histogram) (data from (Harley & Reynolds, 1987)). The SD sequence and start codon for *tnpA* are shown in bold. Mutations introduced into *tnpA/art200* in this work (LS and M1) are indicated in italics. A DNA primer used to map the 5′end of *art200* in *E. coli* (Figure 4.2) is depicted with an asterisk followed by a dashed arrow.
repressed by STnc490, which we renamed art200 (antisense regulator of transposase IS200). Hfq does not play a role in art200-tnpA pairing, although it does repress tnpA expression in the absence of art200. This repression appears to be the result of Hfq binding to the 5’UTR of tnpA immediately upstream of the SD. We also show that the tnpA SD sequence is sequestered in secondary structure and that this inhibits 30S ribosome subunit binding. Finally, we demonstrate that IS200 transposition increases in *E. coli* upon disruption of translational control mechanisms. Implications of these results are considered in the context of tight regulation of IS200 transposition, a possible role for art200 in the control of host gene expression and the potential application of the IS200 sRNA system in synthetic biology/metabolic engineering.

4.2 Results

4.2.1 Characterization of the IS200 antisense RNA gene in *E. coli*

*Salmonella* Typhimurium LT2 contains 6 copies of IS200 and expresses STnc490 (an RNA that is antisense to the transposase RNA) at high levels under standard laboratory growth conditions (Kroger et al, 2012; Sittka et al, 2008). However, the gene encoding this transcript was not previously fully characterized. We show below that the IS200 asRNA is expressed in *E. coli* and characterized components of the gene encoding this transcript. The putative transcription start site for the IS200 asRNA (predicted based on RNA-Seq experiments in *Salmonella* (Kroger et al, 2012)) is shown in Figure 4.1B. We scanned upstream of this position for possible promoter elements and based on consensus sequences for -10 and -35 elements and the optimal spacing between these elements in *E. coli* (Harley & Reynolds, 1987), we defined putative -10 and -35 promoter elements for the IS200 asRNA (Figure 4.1B). We then introduced various mutations into the promoter (Pₐ) of the asRNA (Pₐ-1 to Pₐ-6), which were designed to affect the predicted -10/-35 elements or the surrounding sequences (in the context of a multi-copy plasmid encoding IS200) and performed primer extension analysis (Figure 4.2). The results from our ‘WT’ construct revealed two 5’ ends for the IS200 asRNA (nts 153 and 154), one of which matches the transcription start site previously identified in *Salmonella* (Kroger et al,
2012). Based on the signal intensity it is evident that the IS200 asRNA is abundantly expressed and the two start sites are used at roughly an equal frequency. Variants, including P_{A}-2, P_{A}-3, P_{A}-4 and P_{A}-6, showing the greatest reduction in asRNA expression (Figures 4.2B,C) all had one or more mutations in either the predicted -35 or -10 region, supporting our assignment of these promoter elements. We note that relative to other *E. coli* genes the predicted -35 element of the antisense gene is a better match to the consensus than is the -10 element, the latter of which was almost unrecognizable. In addition to defining fundamental information regarding the IS200 antisense gene in *E. coli*, results from this analysis also provided us with a means of knocking down the levels of the IS200 asRNA in *E. coli*. 
Figure 4.2. Primer extension analysis of art200 in E. coli.

(A) The DNA sequence of the 5’end of the art200 gene plus mutations introduced to deduce the -35/-10 promoter elements are shown. (B,C) Primer extension reactions performed to detect art200 from WT and the indicated mutant forms of tnpA-lacZ are shown. Total RNA was isolated from exponential phase cells grown in rich media (LB) and the primer (5’end labeled with $^{32}$P) shown in Figure 4.1B was used to make cDNA. The cells contained IS200 on a multi-copy plasmid. cDNAs were analyzed on a 10% denaturing polyacrylamide (sequencing) gel. The image in (B) shows a full gel; M = markers, U and C are RNA sequencing reactions. The image in (C) shows only the portion of a gel that includes the art200 primer extension products. For the latter, primer extension was multiplexed to include lpp as a loading control. The mutant $P_{A^{-6}}$ (bold letters) was used in subsequent experiments in this work to knock-down art200 levels in E. coli.
4.2.2 The IS200 antisense RNA inhibits IS200 transposase expression

To measure the impact of the IS200 asRNA on \textit{tnpA} expression we constructed in a multi-copy plasmid an IS200-\textit{lacZ} translational fusion (TLF) in which codon 60 of the \textit{tnpA} gene was fused to codon 10 of the \textit{lacZ} gene (Figure 4.3A). In this construct \textit{tnpA-lacZ} expression was under the control of native IS200 regulatory elements. We also made a version of the TLF plasmid in which the \textit{P}\textsubscript{A}-6 mutations (see Figure 4.2A) had been incorporated. These TLFs were introduced into a \textit{Δlac} strain of \textit{E. coli} (MC4100 derivative) and \textit{tnpA-lacZ} expression was measured using the Miller assay. We show in Figure 4.3B (compare columns 1 and 5) that knocking down IS200 asRNA expression increased \textit{tnpA-lacZ} expression about 13-fold. We also performed primer extension on cells used in the Miller assay and this confirmed that asRNA levels were extremely low in the strain containing the \textit{P}\textsubscript{A}-6 TLF (compare lanes 1 and 5 in Figure 4.3B).

A potential drawback of using the \textit{P}\textsubscript{A}-6 mutations to knock down IS200 asRNA levels was that the \textit{tnpA} RNA sequence is altered by these mutations and this could affect the stability of the \textit{tnpA} transcript and therefore its expression in the Miller assay. Accordingly, we developed an alternative means of knocking down the antisense RNA that did not alter the sequence of the \textit{tnpA} transcript. In this approach we expressed a segment of the \textit{tnpA} mRNA that included a region of the mRNA that was fully complementary to the antisense RNA (nt 45-298 of IS200). Pairing of the two RNAs would potentially promote the degradation of one or both RNA molecules through the action of double-strand specific ribonucleases. This ‘RNA titration’ approach has previously been used in the IS\textit{10} system to decrease levels of the IS\textit{10} encoded asRNA and was also used to knock-down endogenous levels of the MicA sRNA in \textit{E. coli} (Simons & Kleckner, 1983; Udekwu et al, 2005). For our purposes we prepared two titrator constructs differing only in the strength of the promoters used to drive titrator expression (Figure 4.3A; Titrator-high and Titrator-low). We show in Figure 4.3B that titrator RNA expression from both constructs increased \textit{tnpA} expression and the fold increase correlated well with the amount of titrator RNA expressed (compare Miller units in columns 1-3 and the corresponding lanes in the image from primer extension analysis).
Figure 4.3. Impact of art200 and Hfq on \textit{tnpA} expression.

\textbf{(A)} An IS200-\textit{lacZ} translational fusion (TLF) was constructed to measure \textit{tnpA} expression in \textit{E. coli}. Art200 levels were manipulated by: (i) introducing the P\textsubscript{A-6} mutations into the TLF (down-regulated); (ii) co-expressing an art200 titrator RNA (IS200 45-298) with the TLF (down-regulated); or (iii) co-expressing art200 in trans relative to the TLF (up-regulated). \textit{tnpA} expression was entirely under the control of native (IS200) regulatory elements. Titrator RNAs were constitutively expressed from promoters P\textsubscript{tet} (moderate strength) or T7 P\textsubscript{A1} (strong). Trans-art200 was expressed from its native promoter. \textbf{(B)} \(\beta\)-galactosidase and primer extension assays were performed on \textit{E. coli} cells that had been co-transformed with a TLF plasmid (WT or P\textsubscript{A-6}) and a plasmid encoding titrator RNA, trans-art200 or an empty vector control. Cells were grown to mid-log phase in LB supplemented with antibiotics (for plasmid selection)
before being processed for β-galactosidase activity measurements (Miller assay) and primer extension analysis. Bars show the mean β-galactosidase activity from 3 independent experiments (n=9), and error bars indicated standard error on the mean. For primer extension, primers specific for art200, tnpA/titrator and lpp were used. The latter acted as a loading control. (C) tnpA expression was measured as in (B) with isogenic hfq+ and hfq- cells harboring either the WT or PA-6 TLF plasmid. Bars show the mean β-galactosidase activity from 2 independent experiments (n-6) and error bars indicate standard error on the mean.
We also asked if the IS200 antisense RNA could function in trans to repress \textit{tnp}A expression. For this experiment we cloned the antisense RNA gene into a plasmid compatible with our TLF plasmids and co-transformed these plasmids into \textit{E. coli} cells. In the situation where \textit{tnp}A expression was expected to be relatively high because of the P\textsubscript{A}-6 mutations in the TLF (knock-down antisense RNA expression in cis), expression of the antisense RNA in trans reduced \textit{tnp}A-\textit{lacZ} expression about 3.5-fold (compare columns 5 and 8).

Based on the results presented in this section we conclude that the IS200 antisense RNA does function \textit{in vivo} to down-regulate IS200 \textit{tnp}A expression. Accordingly, we have renamed this RNA art200 for \textbf{a}ntisense \textbf{r}egulator of \textbf{t}ransposase IS200.

\textbf{4.2.3 Hfq negatively regulates \textit{tnp}A expression but independent of art200}

Based on previous work in the Tn10 system where we demonstrated that Hfq promotes antisense RNA pairing with the transposase RNA, potentially through restructuring of both RNAs (Ross et al, 2013), we wanted to test the possibility that Hfq might play a similar role in the IS200 system. Towards this end we repeated the experiment described in Figure 4.3B in isogenic \textit{hfq}\textsuperscript{+} and \textit{hfq}\textsuperscript{−} strains of \textit{E. coli}. We show in Figure 4.3C that \textit{tnp}A-\textit{lacZ} expression increased approximately 5-fold in the \textit{hfq}\textsuperscript{−} relative to the \textit{hfq}\textsuperscript{+} strain in the context of the WT TLF (compare columns 1 and 5). This showed that Hfq does repress \textit{tnp}A expression. Additionally, the P\textsubscript{A}-6 and \textit{hfq} mutations acted synergistically to de-repress \textit{tnp}A expression (compare column 1 to 3 and 7), and art200 provided in trans was able to repress expression of the P\textsubscript{A}-6 TLF regardless of Hfq status (compare columns 3 and 4 to columns 7 and 8). Finally, we performed primer extension analysis to measure \textit{tnp}A and art200 levels in \textit{hfq}\textsuperscript{+} and \textit{hfq}\textsuperscript{−} cells and found that \textit{tnp}A levels decreased in the absence of Hfq while art200 levels were unaffected (Figure S4.1). Thus we conclude that Hfq and art200 represent two distinct regulatory mechanisms that down-regulate \textit{tnp}A expression independent of one another.
4.2.4 art200 and \textit{tnpA} mRNA interact \textit{in vitro}

Given that art200 and \textit{tnpA} are complementary over 90 nt it seemed likely that art200 would inhibit \textit{tnpA} expression through complementary base-pairing. However, based on structure probing analysis of art200 and the first 173 nt of \textit{tnpA} mRNA (\textit{tnpA}_{1-173}) along with secondary structure predictions, it is apparent that both RNAs are highly structured and this could limit their ability to pair (Figures S4.2 and S4.3). An alternative possibility is that art200 acts via a protein titration mechanism. We tested for RNA pairing by performing lead(II) acetate (Pb$^{2+}$) footprinting on a mixture of 5'\textsuperscript{32}P-labeled art200 and unlabeled \textit{tnpA}_{1-173}. Both RNAs were generated by \textit{in vitro} transcription and allowed to fold before mixing. Pairing would convert single to double stranded regions and consequently there would be a loss of Pb$^{2+}$ reactivity at these positions. We show in Figure 4.4A that a cluster of residues (marked with a green asterisk) in the upper portion of the predicted stem-loop of art200 exhibited reduced Pb$^{2+}$ reactivity upon addition of \textit{tnpA}_{1-173}. Note that most of these residues were in parts of art200 predicted by our model to be single stranded. The complementary nucleotides in \textit{tnpA}_{1-173} include positions -23 to -62.

We also performed the complementary experiment with 5'\textsuperscript{32}P-labeled \textit{tnpA}_{1-173}, but in this case used RNases (A, T1, and V1) as structure probes (Figure 4.4B). Both RNase A and T1 are single strand specific and accordingly reduced reactivity with these enzymes in the presence of unlabeled art200 would provide evidence of base-pairing. Comparison of lanes 7 and 8 (RNase A) and lanes 12 and 13 (RNase T1) revealed two areas containing the most prominent reactivity decreases including residues -60 to -23 and -7 to +11 (indicated by red asterisks). The former region encompasses the upper stem-loop of \textit{tnpA}_{1-173}, thus supporting results from the Pb$^{2+}$ footprinting that were consistent with the upper stem-loop region of art200 participating in base-pairing with the upper stem-loop of \textit{tnpA}_{1-173}. Also consistent with this interpretation, there were several examples of nucleotides in this region showing increased reactivity to RNase V1 a double-strand specific ribonuclease (blue asterisks). A summary of the footprinting data is presented in Figure 4.4C (Pb$^{2+}$, green; RNase A/T1, red; RNase V1, blue).
Figure 4.4. Pb$^{2+}$ and RNase footprinting of art200, *tnpA*$_{1-173}$ and an art200-*tnpA*$_{1-173}$ complex.

(A) 5$^3$P-labeled art200 (69 nM) was incubated in the absence or presence of increasing concentrations of *tnpA*$_{1-173}$ (69, 128, 276, 460, or 1380 nM) before limited treatment with Pb$^{2+}$. Note that each RNA was denatured and allowed to fold before mixing. Positions that were most strongly protected from Pb$^{2+}$ cleavage in a *tnpA*-concentration dependent manner are indicated with a green asterisk. UT is untreated art200 RNA and G is an RNase T1 sequencing lane. (B) 5$^3$P-labeled *tnpA*$_{1-173}$ (40 nM) was incubated with wild-type and mutant variants (LS’ and M1 – see Figure 4.1B) of art200 (600 nM) or folding buffer (-) before treatment with RNase A, T1, or V1. *tnpA*$_{1-173}$ and art200 RNA were denatured and allowed to fold independently before mixing, except for a control reaction with WT art200 where RNAs were mixed, denatured and allowed to fold together (FT; lanes 6, 11, 16, and 21). Nucleotide numbering is relative to the AUG start codon in *tnpA*. Nucleotides that were most strongly protected from single-strand specific RNase (A/T1) in the presence of art200 are indicated with a red asterisk, and positions that showed an increased sensitivity to RNase V1 (double-strand specific) are indicated with a blue asterisk. (C) Structural constraints derived from footprinting were input into mFold to produce structures for art200 and *tnpA*$_{1-173}$ (see also Figures S4.2 and S4.3). Residues in art200 that showed either weak (green circle) or strong (green circle plus asterisk) decreases in Pb$^{2+}$ reactivity upon mixing with *tnpA*$_{1-173}$ are highlighted. Residues in *tnpA*$_{1-173}$ that showed strong (red circles) decreases in RNase A or T1, or strong increases (blue circles) in V1 reactivity upon art200 addition are highlighted. Two residues (-44 and -47) showed increased V1 sensitivity and decreased A/T1 sensitivity (blue-red circles). Nucleotide changes present in M1 and M1’ versions of art200 and *tnpA*$_{1-173}$, respectively, are shown in bold.
The terminal loops of art200 and tnpA_{1-173} include 4 and 6 unpaired residues respectively, and have the highest G-C content of any of the single stranded regions in the two RNA molecules. This led us to predict that art200-tnpA_{1-173} pairing might initiate with a kissing loop interaction involving these two loops. Accordingly, we mutated three residues in the terminal loop of art200 (art200^{M1}) and asked if this form of art200 could still pair with tnpA_{1-173} using RNase footprinting. For all of the residues that showed decreased RNase A or T1 reactivity in the presence of art200^{WT} (red asterisks), we observed reduced protection in the presence of art200^{M1}. Similarly, all of the residues that showed increased V1 reactivity in the presence of art200^{WT} showed decreased reactivity in the presence of art200^{M1}. We also introduced mutations to the lower stem region of art200 (nts 78-84, LS’) and observed an intermediate effect on pairing relative to WT and M1 suggesting this region is less important for pairing.

The above results show that art200 and tnpA do indeed interact in vitro; however pairing is limited to loosely structured regions of both RNAs. In particular, the terminal loop region of each RNA is important for pairing, which may indicate that pairing initiates with these sequences through a kissing loop interaction.

4.2.5 Base-pairing between art200 and tnpA blocks 30S ribosome binding in vitro and inhibits transposase expression in vivo

Based on our Pb^{2+} and RNase footprinting experiments we thought it likely that art200 pairing with the 5’UTR of tnpA would inhibit translation initiation. We tested this possibility by performing toeprinting analysis. In this assay purified 30S ribosomal subunit plus initiator tRNA (fMet-tRNA) was mixed with tnpA_{1-173} either in the presence or absence of art200. Primer extension with a 5’^{32}P-labeled primer complementary to nucleotides +51 - +70 on tnpA_{1-173} was then performed and reactions were analyzed on a sequencing gel. Typically the 30S ribosome leaves a footprint of ~30 nucleotides spanning the SD sequence and first 5 codons such that a strong stop is produced in the primer extension reaction about 15 nt downstream of the start codon. We show that in the absence of art200, a relatively weak stop signal was observed at position +16 when 30S ribosome and fMet-tRNA were incubated with tnpA_{1-173} (Figure 4.5A, lane 6); the weak toeprint signal is consistent with previous work suggesting that the tnpA SD sequence is
sequestered in a secondary structure element (Beuzon et al, 1999) (see also Figure S4.3). Addition of art200 (15:1 molar excess) inhibited formation of the toeprint signal by approximately 95% (lane 7 and Figure 4.5B). However, when we used art200\textsuperscript{M1} instead of art200\textsuperscript{WT} inhibition of the toeprint signal was greatly reduced to approximately 50%. Given the evidence presented in Figure 4.4 that art200\textsuperscript{M1} fails to base-pair with \textit{tnpA}, the inability of art200\textsuperscript{M1} to suppress the toeprint signal to the same degree as art200\textsuperscript{WT} is consistent with art200 inhibiting 30S ribosome binding through a base-pairing interaction with \textit{tnpA}. We also show that when \textit{tnpA} is mutated to restore complementarity with the terminal loop region of art200\textsuperscript{M1}, ribosome binding is blocked by art200\textsuperscript{M1} but not art200\textsuperscript{WT} (compare lanes 12-14 in Figure 4.5A).

Finally, the toeprint analysis also provided further details of the \textit{tnpA}-art200 pairing interaction. In all of the reactions that included a form of art200 that was fully complementary to \textit{tnpA} (lanes 5, 7, 11 and 13) there were a series of prominent primer extension pauses upstream of the SD. These strong pauses can be explained by art200 pairing with \textit{tnpA} and thus the experiment reveals that position -25 in \textit{tnpA} defines a ‘downstream’ boundary of antisense pairing. This fits well with our structure probe data, which was consistent with position -23 being the downstream boundary.

We also looked at the impact of terminal loop mutations (art200\textsuperscript{M1} and \textit{tnpA}\textsuperscript{M1'}) on \textit{tnpA-lacZ} expression \textit{in vivo}. We show in Figure 4.6 that when art200 was provided in trans (in the \textit{P\textsubscript{A}-6} TLF background) strong repression of \textit{tnpA} was only achieved when the terminal loops of art200 and \textit{tnpA} were perfectly complementary (compare columns 2 and 3). Also, trans-art200\textsuperscript{M1} was capable of repressing expression of \textit{tnpA}\textsuperscript{PA-6/M1’-lacZ} but not \textit{tnpA}\textsuperscript{PA-6-lacZ} (compare columns 3 and 6). Primer extension analysis on RNA prepared from the strains in Figure 6 showed that both forms of trans-art200 were expressed at similar levels in these experiments (Figure 4.6, lower panel).

Taken together the results from experiments in Figures 4.5 and 4.6 show that despite 90 nt of perfect complementarity, the primary determinant for antisense
Figure 4.5. Impact of antisense pairing on ribosome binding to \( tnpA_{1-173} \) \textit{in vitro}. 

(A) 30S ribosome binding to WT and M1’ \( tnpA_{1-173} \) was measured in a toeprint assay. Where indicated, WT or M1 art200 (3 μM) was added to \( tnpA \) RNA (200 nM) prior to
addition of the 30S ribosomal subunit and initiator tRNA. Strong pauses in reverse transcription (G_{+15}/G_{+16}) produced upon incubating the above mix with reverse transcriptase, dNTPS and a 5'\textsuperscript{32}P-labeled DNA primer (anneals downstream of the \textit{tnpA} start codon) define the toeprint signal. Positions of prominent art200-dependent pauses in reverse transcription that occur independent of 30S ribosome addition are also indicated. G, A, and C are sequencing lanes and nucleotide numbering is relative to the start codon of \textit{tnpA}. (B) Toeprint signal band intensities from (A) were quantified. The toeprint signal for \textit{tnpA}\textsuperscript{WT} in the absence of art200 was set at 100%.
Figure 4.6. Impact of terminal loop mutations on translational repression in vivo.

A plasmid encoding tnpA<sup>PA-6</sup>-lacZ or tnpA<sup>PA-6/M1</sup>-lacZ was co-transformed into DBH323 with a compatible plasmid expressing art200 (WT or M1) in trans to tnpA-lacZ or an empty vector control. β-galactosidase activity was measured in cells grown to mid-exponential phase in LB media. Bars show the mean expression from two independent experiments and error bars indicate standard error on the mean (n=6). The bottom panel shows primer extension analysis using RNA extracted from cells grown in the Miller assay. lpp was analyzed as a loading control.
repression in the IS200 system is complementarity between the upper stem-loop regions of tnpA and art200 and that base-pairing between residues in the terminal loops is critically important for antisense repression. Although we have not investigated the effect of other mutations in single-stranded regions of either RNA (e.g. nt 62 to 65 in art200 and -53 to -56 in tnpA) it seems likely that pairing initiates with the 3 G/C base-pairs affected by the M1 mutation and then propagates roughly half-way down the respective stems. An initial kissing-loop interaction has been shown in many other antisense systems to be important for pairing (Brantl, 2002; Franch et al, 1999; Hjalt & Wagner, 1992). Further pairing might be inhibited by the absence of bulges in the lower portions of the respective stems, as such discontinuities in intramolecular base-pairing have been shown in other studies to be important in destabilizing stem structures and allowing intermolecular base-pairing (Brantl, 2007; Hjalt & Wagner, 1995; Kittle et al, 1989).

4.2.6 tnpA translation is also repressed by mRNA secondary structure

Previous work in the IS200 system revealed that deleting the 5’ portion of tnpA mRNA (nts -32 to -103) resulted in a ~10-fold increase in tnpA-lacZ expression (Beuzon et al, 1999). The authors from this study concluded that the increased expression resulted from the loss of an inhibitory stem-loop structure; however, their deletion also removed half of art200. To determine if RNA secondary structure plays an important role in inhibiting tnpA expression, we introduced mutations to the lower stem (nts -69 to -75) and evaluated the impact of these mutations on ribosome binding in vitro and on tnpA expression in vivo. We show in the toeprinting assay in Figure 4.7A that tnpA<sub>11-173</sub> with the lower stem mutations (tnpA<sub>LS</sub>) gave a much higher toeprint signal (20-fold increase) than tnpA<sub>WT</sub> (compare lanes 6 and 11; also see Figure 4.7B). This indicates that nucleotides comprising the lower stem are important determinants for tnpA translation (as previously suggested). We also asked if art200 could still repress ribosome binding in the tnpA<sub>LS</sub> background. Both art200<sub>WT</sub> and art200<sub>LS’</sub> strongly repressed ribosome binding (compare lanes 12 and 13 with lane 11) but a mutant form of art200 (art200<sub>M1</sub>) lacking full terminal loop complementarity with tnpA<sub>LS</sub> failed to fully block ribosome binding.
Figure 4.7. Impact of *tnpA* lower stem (LS) mutations on transposase expression.

(A) The toeprinting assay was performed on *tnpA* RNA (WT and LS; 200 nM) in the presence or absence of art200 (WT, LS’, or M1; 3µM) as described in Figure 4.5. The toeprint signal spans nucleotides 15-17. C, U, A, and G are sequencing lanes. (B) Toeprint signal band intensities (G+15, A+16, and G+17) from (A) were quantified. The toeprint signal for *tnpA*<sup>WT</sup> in the absence of art200 was set at 100%. (C) A plasmid encoding *tnpA*<sup>WT</sup>-lacZ or *tnpA*<sup>LS</sup>-lacZ was co-transformed into DBH323 with a compatible plasmid expressing art200 titrator RNA (WT or M1) or an empty vector control. β-galactosidase activity was measured in cells grown to mid-exponential phase in LB media. Bars show the mean expression from two independent experiments and error bars indicate standard error on the mean (n=6).
(lane 14). These results indicate that the nucleotides comprising the lower stem of \textit{tnpA} are not critical for antisense repression and thus contribute to a distinct mode of \textit{tnpA} translational regulation.

We also determined the impact of the LS mutations on \textit{tnpA} expression \textit{in vivo}. Consistent with the toeprinting assay, the LS mutations increased \textit{tnpA}-\textit{lacZ} expression 50-fold relative to that observed for WT \textit{tnpA}-\textit{lacZ} (compare columns 1 and 3 in Figure 7C). Titration of art200\textsubscript{LS} in this system with the high copy titrator further increased expression 4-fold (column 4) indicating that the two regulatory systems can act independent of each other to repress \textit{tnpA} expression.

We therefore conclude that in addition to a cis-encoded sRNA, translation of the IS200 transposase is strongly repressed by an mRNA secondary structure that can directly sequester the SD.

4.2.7 Hfq blocks ribosome binding to \textit{tnpA} \textit{in vitro}

Although Hfq is not required for antisense pairing, \textit{tnpA}-\textit{lacZ} expression increased 5-fold in an \textit{hfq}\textsuperscript{−} versus \textit{hfq}\textsuperscript{+} strain of \textit{E. coli}. In addition, we have shown that this up-regulation in the absence of Hfq did not require the production of art200 (Figure 4.3C). This indicates that Hfq represses transposase expression in an antisense-independent manner. Additionally, \textit{tnpA} levels do not increase in \textit{hfq}\textsuperscript{−} which indicates Hfq acts at the level of \textit{tnpA} translation (Figure S4.1). We have recently shown that in the IS10 system, Hfq binding to the ribosome binding site of transposase mRNA was sufficient for repressing translation initiation (Ellis et al, 2015a). We therefore considered the possibility that Hfq might be acting directly on \textit{tnpA} to inhibit translation.

We show in Figure 4.8A and B that Hfq inhibited formation of the \textit{tnpA}_{1-173} toeprint in a concentration dependent manner (see also Figure S4.4). At a 1:1 molar ratio of Hfq:\textit{tnpA}, the toeprint signal was reduced 40-50\% compared to no Hfq addition and at a 4:1 ratio of Hfq:\textit{tnpA} the toeprint signal was reduced 80\%. Thus, Hfq can block 30S ribosomal subunit binding to \textit{tnpA}_{1-173} \textit{in vitro} independent of an sRNA. We note that the
Figure 4.8. Hfq inhibits 30S ribosomal subunit binding to tnpA and binds upstream of the SD.

(A) Toeprint assay showing the effect of Hfq on 30S ribosomal subunit to tnpA. Hfq (100–400 nM; hexamer concentration) was added to tnpA (200 nM) prior to addition of 30S ribosomal subunit and initiator tRNA. A section of the gel image including the toeprint signal is shown. (B) The percent inhibition of toeprint signal upon incubating Hfq with tnpA or an mRNA that does not interact with Hfq (usg, (Beisel et al, 2012)) is shown; the usg data comes from (Ellis et al, 2015a). For both mRNAs the toeprint signal in the absence of Hfq was set at 100%. Experiment A refers to (A) of this figure while Experiment B refers to Figure S4.4. (C) Hydroxyl radical footprinting experiment with 5'32P-labeled tnpA1-173 (68 nM) and the indicated concentrations of Hfq. Subsequent to mixing tnpA and Hfq, limited RNA cleavage by hydroxyl radical treatment was carried out as previously described (Ross et al, 2013). UT is untreated RNA and G is an RNA cleavage ladder produced by RNase T1. The Hfq footprint between residues -17 and -33 defined an Hfq binding site in tnpA and the position of this site is highlighted (grey circles) in our model for tnpA1-173; the tnpA SD and start codon are in bold.
strength of the Hfq block on ribosome binding in the IS200 system is weaker than previously seen with the IS10 transposase mRNA (RNA-IN) but slightly stronger than observed with a control mRNA, usg (Figure 4.8B) (Ellis et al, 2015a). For example, when Hfq is limiting (1:2 ratio of Hfq:mRNA), the toeprint signal was reduced 30% for tnpA and only 12% for usg mRNA. However, at higher concentrations of Hfq the toeprint signal was reduced a comparable amount for both tnpA and usg.

We further analyzed the Hfq-tnpA interaction by performing hydroxyl radical footprinting on 5'\textsuperscript{32}P-labeled tnpA\textsubscript{1-173} mixed with various concentrations of Hfq (Figure 4.8C). The results of the footprinting were consistent with Hfq binding tnpA in an interval extending from position -33 to -17. Together, the above results suggest that Hfq binding immediately upstream of the tnpA SD sequence represses tnpA translation by preventing ribosome binding.

4.2.8 IS200 transposition is limited by translational control

Typically for bacterial transposons, transposition frequency correlates strongly with transposase expression (Kleckner, 1990; Morisato et al, 1983; Nagy & Chandler, 2004). We measured IS200 transposition by constructing a mini-IS200 element (IS200-kan) and using this marked element in mating out experiments (see Figure S4.5 for schematic of the mating out assay). IS200 transposase was provided in trans from a plasmid in which the tnpA gene was under the control of different regulatory elements (Figure 4.9A).

We did not detect transposition events when tnpA expression was under the control of the fully native regulatory elements. We did detect transposition events when tnpA was fused to P\textsubscript{BAD} and SD\textsubscript{BAD24} and arabinose (0.2%) was present during growth (construct i). Notably, the number of events was considerably higher than in a control where the tnpA gene contained a mutation in the catalytic tyrosine (construct ii). We confirmed that these were authentic transposition events by mapping two independent hops from construct (i) using ST-PCR (Chun et al, 1997; Ross et al, 2014) (Figure S4.6A). We therefore conclude that the IS200 TnpA protein from Salmonella is active for transposition in E. coli.
When we replaced SD$_{BAD24}$ with the native 5’UTR (construct $iii$), the transposition frequency dropped considerably; a single transposition event was observed in one of three experiments. This construct produced a large amount of $tnpA$ mRNA suggesting that translational control strongly limits transposition (Figure 4.9B). We then introduced mutations into the native 5’UTR to disrupt the $tnpA$ stem-loop (construct $iv$). This increased the frequency of transposition events, although the occurrence of these events was still sporadic.

We next measured transposition from construct ($iv$) in the presence of the art200 titrator plasmid or a vector control. Our expectation was that disrupting two regulatory pathways (mRNA structure and antisense control) would further increase transposition. In the presence of the titrator plasmid, 8/10 donor isolates produced measurable transposition while only 5/10 donors produced hops in the presence of the vector control (Figure 4.9C). This coincided with a 25-fold increase in the median value of transposition when art200 was depleted. Together this data shows that (1) TnpA expression is in fact limiting for IS$_{200}$ transposition, and (2) disrupting translational regulation (mRNA secondary structure and antisense control) of $tnpA$ leads to an increase in IS$_{200}$ transposition.
Figure 4.9. IS200 transposition assays.

(A) IS200 transposition frequency was measured using the conjugal mating out assay. Briefly, *E. coli* (F<sup>+</sup>; DBH291) containing a single chromosomal copy of a marked IS200 element (mini IS200-kan) was transformed with a plasmid expressing TnpA under the control of various regulatory elements, including the P<sub>BAD</sub> promoter, the 5′UTR from pBAD24 (includes an optimized SD) and the IS200 5′UTR (constructs i-iv). These donor cells were grown in the presence of arabinose (0.2%) to induce *tnpA* transcription, mixed with an F<sup>−</sup> recipient strain (DBH13) and then the mating mixes were plated on selective media for measuring mating efficiency (exconjugants) and transposition events (hops). Transposition frequency is the ratio of hop to exconjugant colonies. Transposition frequencies for individual donor clones are presented in a stripchart for each TnpA construct: grey bars show the median transposition frequency for one (constructs i and ii) or three (constructs iii and iv) independent experiments. Clones that did yield hops and were analyzed by Southern blot analysis are indicated (a-c). LE = left end (bp 1-163), RE
= right end (bp 566-707), and kan$^R$ = kanamycin resistance gene. (B) Primer extension analysis of DBH291 donor cells transformed with construct (iii) and grown to mid-log phase in the presence of arabinose. Primer extensions were multiplexed to detect tnpA, art200, and lpp (loading control). (C) Mating out assay with conor strains containing construct (iv) and either the low expression art200 titrator plasmid or an empty vector control. Grey bars show the median transposition frequency for each donor strain from three independent experiments; d, e, f, and g are hop colonies subjected to Southern blot analysis (Figure S4.6). In (A) and (C) the transposition frequency for donor clones that did not produce hop colonies was set at 1 x 10$^{-8}$.
4.3 Discussion

4.3.1 Translation of the IS200 transposase is repressed by a cis-encoded sRNA, Hfq and RNA secondary structure

IS200 is a very unusual transposable element in that it is widespread in Eubacteria and in some species has attained a very high copy-number (see below), yet its ability to transpose is exceedingly poor. This correlates with very weak expression of the IS200 transposase protein. In the current work we have expanded our understanding of how IS200 transposase expression is suppressed to include two new levels of post-transcriptional regulation and further characterization of a predicted cis-regulatory element. First, we show that the recently identified sRNA art200 (previously STnc490 (Sittka et al, 2008)) encoded opposite the transposase 5’UTR represses transposase translation by base-pairing with tnpA mRNA and blocking 30S ribosome binding. Additionally, we expand on previous work that suggested RNA secondary structure in the 5’UTR of tnpA inhibits translation by sequestering the SD in a stable stem loop structure (Beuzon et al, 1999). Finally, we show that the chaperone protein Hfq is also a negative regulator of tnpA expression. Footprinting revealed that Hfq binds immediately upstream to the tnpA SD raising the possibility that Hfq could block 30S subunit binding to tnpA. Support for this came from toeprinting studies where at low concentrations of Hfq (100 nM) ribosome binding to tnpA was reduced 30%. It is not clear at this point if this reduction is significant as the level of toeprint inhibition was only marginally higher than that detected in a control reaction; 15% with usg mRNA. By comparison in another system (cirA mRNA) where Hfq was reported to directly interfere with 30S subunit binding a similarly small reduction in toeprint signal (20%) was reported at low Hfq concentrations (Salvail et al, 2013). Based on this data we suggest that this moderate effect on ribosome binding could account for at least a portion of the 5-fold repression Hfq has on tnpA-lacZ expression in vivo. Notably, this represents the first example of a bacterial ‘host’ protein suppressing IS200 (Beuzon et al, 2004) and the second example of Hfq directly repressing translation of a transposase protein (Ellis et al, 2015a).

An interesting aspect of the regulatory mechanisms described here is that all three are capable of acting independently to interfere with 30S subunit binding to tnpA. This
conclusion comes from the following observations: (1) art200 repressed \textit{tnpA} expression and ribosome binding in the absence of Hfq (Figures 4.3C and 4.5); (2) art200 suppressed 30S subunit binding to \textit{tnpA} under conditions where the inhibitory stem-loop structure is destabilized by mutations in \textit{tnpA} (Figure 4.7A); and (3) the effect of disrupting \textit{hfq} on \textit{tnpA}^{WT} expression was almost 20-fold less than inhibiting formation of the stem-loop structure (compare Figure 4.3C and 4.7C). If Hfq acted to stabilize the stem-loop structure one might have expected \textit{tnpA} expression to be comparable in \textit{hfq}^{-} and \textit{tnpA}^{LS} situations.

What might be the explanation for this level of functional redundancy? IS200 \textit{tnpA} contains an almost perfect Shine-Dalgarno sequence (\textit{tnpA}, AAGGGGGU; \textit{E. coli} consensus, AAGGAGGU) (Shine & Dalgarno, 1975). However, this sequence is sequestered in secondary structure (this work and (Beuzon et al, 1999)). Interestingly, upstream of the SD there is a single-stranded C/A-rich sequence (nts -26 to -21) that potentially could act as a translational enhancer. Such sequences can provide an initial toehold for the 30S ribosomal subunit through a direct interaction between the S1 protein component of the 30S complex and the C/A-rich RNA sequence (Komarova et al, 2002; Mogridge & Greenblatt, 1998; Ringquist et al, 1995; Sengupta et al, 2001). S1 could tether the ribosome to \textit{tnpA} and expose the downstream SD sequence for 30S subunit binding by altering the local RNA structure (Bear et al, 1976; Qu et al, 2012; Rajkowitsch & Schroeder, 2007; Studer & Joseph, 2006). As we have shown that art200 pairs with the C/A-rich containing portion of the \textit{tnpA} transcript and Hfq binds this same region, it is possible that both art200 and Hfq repress the function of this putative translational enhancer sequence by sterically occluding S1 binding. The combination of sequestration of the SD and interference of translational enhancer function would be expected to provide a very strong block (synergistic or at least additive) to translation, which we observed here. We have some evidence of the C/A-rich region playing a regulatory role in translation as mutations in this region reduced \textit{tnpA} expression almost 200-fold (Figure S4.7), although as we were unable to measure steady-state \textit{tnpA} RNA levels because of the extremely low abundance of this transcript we cannot rule out the possibility that this decrease resulted from the mutations destabilizing the transcript.
Notably, there are several examples in the literature of sRNAs interfering with translational enhancer function. The sRNA GcvB represses initiation of translation for multiple mRNA transcripts by pairing with C/A-rich translational enhancers (Sharma et al., 2007; Yang et al., 2014). In addition, other sequences upstream of the SD have been shown to influence 30S subunit binding. In the case of the *tisAB* transcript, which has its SD sequence sequestered in a highly structured region, a genetic element distinct from a C/A-rich translational enhancer was shown to provide a ‘standby’ site for 30S binding. It was inferred that 30S binding to this sequence opened up the downstream structure for subsequent 30S binding to the SD. The sRNA IstR-1 acts as a negative regulator of translation in this system by competing with the 30S subunit for the standby site (Darfeuille et al., 2007). Art200 and/or Hfq could act in a similar manner in the IS200 system (Figure 4.10, (i)).

Finally, it is also possible that art200 and/or Hfq exert their negative regulatory effects on the IS200 system by binding close enough to the SD to directly block 30S subunit binding. It has been reported that the maximal ribosome-binding region can include nucleotides as far as 39 residues upstream of the start codon (Huttenhofer & Noller, 1994) and the art200 pairing site and the Hfq binding site fall within this window. If this latter mechanism were in play in the IS200 system, then all three negative regulatory systems would be acting at the same step in translation and accordingly the reason for this level of redundancy would be less clear. Although one possibility could be that there are some circumstances where SD sequestration would be suboptimal. For example, under conditions where transcription rates are reduced it is possible that the anti-SD sequence in *tnpA* could pair with an alternative sequence to the SD (one such possible structure is shown in Figure S4.8). In this case art200 and/or Hfq could provide important back-up functions for limiting 30S binding to *tnpA* (Figure 4.10, (ii)).

Given the current work, it is not surprising that IS200 transposition is exceptionally rare. In addition to weak transcription of the transposase gene, *tnpA* translation is suppressed by three independent mechanisms. As we have shown that TnpA expression is in fact limiting for transposition (Figure 4.9), we speculate that translation initiation represents the main point of regulation for IS200 transposition.
Figure 4.10. Model for translational repression of IS200 tnpA.

30S ribosomal subunit binding to tnpA is inhibited by art200 (purple) and Hfq (green) as well as RNA secondary structure which sequesters the Shine-Dalgarno sequence (SD, light blue line). Art200 and Hfq may act to (i) block ribosomal protein S1 binding to a translational enhancer or (ii) simply prevent the 30S-SD interaction. An alternative secondary structure of tnpA where the anti-Shine-Dalgarno (anti-SD, red line) is not paired to the SD is derived from secondary structure predictions (Figure S4.8). See discussion for more details.
4.3.2 Might art200 function as both a cis and trans acting sRNA?

We have previously shown that Hfq represses IS10 transposase expression by facilitating the pairing between transposase mRNA (RNA-IN) and a cis-encoded sRNA (RNA-OUT) (Ross et al, 2013) and this led us to ask if similar regulation would occur in the IS200 system with art200. However, the current work shows that Hfq is not required for art200-mediated repression of tnpA expression. There are a large number of Hfq-binding RNAs in vivo and we and others have provided evidence that Hfq is in fact limiting for RNA binding (Ellis et al, 2015a; Hussein & Lim, 2011; Moon & Gottesman, 2011; Papenfort et al, 2009). Since Hfq binding in vivo must therefore be selective, it seems likely that the Hfq-art200 interaction is biologically important (Miyakoshi et al, 2015), although for gram positive bacteria only a subset of Hfq-binding sRNAs seem to rely on Hfq for stability and/or riboregulation (Christiansen et al, 2006; Dambach et al, 2013; Jousselin et al, 2009; Nielsen et al, 2010; Sievers et al, 2014). It is possible that art200 also is a trans-acting sRNA, and that this secondary function requires Hfq.

In addition to its Hfq binding properties, art200 expression increases during stationary phase and under conditions that induce the Salmonella pathogenicity islands (Sittka et al, 2008). There is no a priori expectation that expression of an RNA involved in repressing transposition would fluctuate in response to external stimuli or growth phase. Art200 is also expressed at a level far greater than that required to repress the poorly expressed tnpA mRNA (see Figure 4.3B).

One paradox of IS200 elements is that while these transposons are essentially dormant many genomes containing IS200 elements have multiple copies. For example, natural isolates of Salmonella and Shigella contain up to 25 and 4 copies of IS200, respectively (Gibert et al, 1990) and the Y. pestis 6/69M genome contains at least 30 copies of the closely related IS1541 (Odaert et al, 1998). In fact, a BLAST search for IS200 elements in Salmonella revealed an average of 9.6 (n=33) copies of IS200 per genome, while a similar search in Yersinia averaged 39.8 (n=30) copies per genome. In contrast, E. coli contains an average of 2.9 (n=31) IS200 elements per genome. The high copy-number of IS200 elements in certain species may simply reflect host-specific adaptation by the transposon (Siguier et al, 2014). Alternatively, IS200 might have been
domesticated by certain host bacteria, in which case IS200 expansion (and utilization of art200 as a trans regulator) could be a response to selective pressure. There are several examples of Hfq integrating horizontally acquired genes into host regulatory networks (Papenfort et al, 2012; Papenfort & Vogel, 2010; Pfeiffer et al, 2007; Shakhnovich et al, 2009; Sittka et al, 2008; Vogt & Raivio, 2014) and art200 may represent one such case.

4.3.3 IS200 5’UTR as a platform for designing novel riboregulators of translation initiation

A major goal in the field of synthetic biology is to create tightly controlled gene regulatory networks to coordinate the expression of a range of desired protein products. The ultimate goal of this field is to produce microorganisms capable of producing biomaterials, pharmaceuticals, and biofuels and acting as biosensors for a range of applications (Khalil & Collins, 2010). Since these biosynthetic pathways must be tightly regulated yet easily manipulated, a great deal of work has been done to design riboregulators of transcription and translation. As it is advantageous to adapt naturally occurring regulators rather than de novo design, well-studied systems such as the pT181 transcriptional attenuator and IS10 antisense system have been modified and combined for synthetic biology applications (Chappell et al, 2013; Liu et al, 2012; Mutalik et al, 2012; Qi et al, 2012a; Qi et al, 2012b; Takahashi & Lucks, 2013).

We propose that the IS200 5’UTR will serve as a convenient platform for modular design of orthogonal regulators of protein synthesis. First, we show here that the cis-encoded antisense system can be easily re-programmed by altering 3nt in the terminal loop region of each RNA. We have not investigated the impact of more extensive changes but predict that this could provide greater specificity. Additionally, our work shows that antisense regulation can be exploited for negative regulation (i.e. providing art200 in trans) or positive regulation (i.e. eliminating art200 through titration). In principle, the tnpA 5’UTR could be fused to a gene of interest and translation of this downstream gene could be modified by an art200 derivative provided in trans. Translation could be further regulated by selectively disrupting the secondary structure that naturally occludes the SD sequence on tnpA. An RNA which base-pairs with the linear region of tnpA immediately 5’ to the lower-stem as well as the ‘anti-SD’ sequence could reduce secondary structure
in a manner analogous to the LS mutation we described here. This synthetic RNA would be similar to the recently described ‘trigger RNA’ which can activate expression of *de novo* designed toehold switches (Green et al, 2014).

### 4.4 Materials and Methods

#### 4.4.1 Bacterial strains, plasmids and oligonucleotides

All Miller assays and related RNA analyses were performed in *E. coli* K-12 derivatives DBH323 (MC4100 Δ*recA774::kan*) or DBH326 (DBH323 *hfq-1::cat*) (Ross et al, 2014). *Salmonella enterica* subsp. *enterica* serovar Typhimurium str. LT2 was used as a source of IS200. For mating out experiments, DBH33 was lysogenized with λDBH881 to create the donor strain DBH291 (DBH33 Mini IS200-*kan*) and DBH13 was used as the recipient strain (see Supplemental Materials and Methods for details of strain construction). DH5α was used for routine cloning and plasmid propagation. Strains and plasmids used in the main text are listed in Table 4.1; all other plasmids and oligonucleotides are listed in Tables S4.1 and S4.2 respectively.

The IS200-*lacZ* translational fusion (TLF; pDH861) and mutant derivatives consist of the first 323 nt of IS200 fused to codon 10 of the *lacZ* gene cloned into pGEM-T easy (Promega). The art200 titrator plasmids (pDH898 and pDH899; Figure 4.3A) consist of nt 45-298 of IS200 (no cis-art200) transcribed from either the T7 phage *P*~A1~ or *P*Tet. The *sgrS* transcriptional terminator was inserted immediately downstream of IS200, and the entire construct was cloned into pACYC184. The plasmids expressing art200 in trans (pDH902 and pDH912) consist of nt 45-298 of IS200 cloned into pACYC184. Transposase expression in mating out experiments (pDH857, pDH860, pDH896 and pDH897) was from pBAD24 derivatives (Guzman et al, 1995) where TnpA was expressed from native or exogenous regulatory elements. Further details of constructing these plasmids are provided in Supplemental Material.
<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBH13</td>
<td>HB101 [F- leu- pro-]; Sm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Recipient strain for mating out experiments</td>
</tr>
<tr>
<td>DBH33</td>
<td>NK5830 [recA- arg- ΔlacproXIII nalR rifR/ F' lacpro+]</td>
<td>Parent strain for mating out donor</td>
</tr>
<tr>
<td>DBH291</td>
<td>DBH33 λDBH881 (Mini IS200-kan); Kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Donor strain for mating out experiments</td>
</tr>
<tr>
<td>DBH323</td>
<td>MC4100 ΔrecA774::kan; Sm&lt;sup&gt;R&lt;/sup&gt;Kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Miller assays (hfq&lt;sup&gt;+&lt;/sup&gt;) (Ross et al, 2014)</td>
</tr>
<tr>
<td>DBH326</td>
<td>DBH323 hfq-1::cat; Sm&lt;sup&gt;R&lt;/sup&gt;Kan&lt;sup&gt;R&lt;/sup&gt;Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Miller assays (hfq&lt;sup&gt;-&lt;/sup&gt;) (Ross et al, 2014)</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pDH857</td>
<td>pBAD24-SD&lt;sub&gt;BAD24&lt;/sub&gt;-tnp&lt;sub&gt;W&lt;/sub&gt;T; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>TnpA expression for mating out, pBAD24 regulatory elements</td>
</tr>
<tr>
<td>pDH860</td>
<td>pBAD24-SD&lt;sub&gt;BAD24&lt;/sub&gt;-tnp&lt;sub&gt;Y&lt;/sub&gt;125F; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>TnpA&lt;sup&gt;Y125F&lt;/sup&gt; (cat. dead) expression for mating out, pBAD24 regulatory elements</td>
</tr>
<tr>
<td>pDH861</td>
<td>pGEM-T Easy derived, tnpA-lacZ TLF; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>TLF with disrupted stem-loop structure</td>
</tr>
<tr>
<td>pDH862</td>
<td>pDH861 with lower-stem mutations, tnpA&lt;sub&gt;L&lt;/sub&gt;S- lacZ; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>tnpA&lt;sub&gt;L&lt;/sub&gt;S-lacZ, no cis-art200</td>
</tr>
<tr>
<td>pDH880</td>
<td>pDH861 with P&lt;sub&gt;A&lt;/sub&gt;-6 mutation; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>TnpA expression for mating out, WT IS200 5’UTR</td>
</tr>
<tr>
<td>pDH896</td>
<td>pBAD24-tnp&lt;sub&gt;W&lt;/sub&gt;T; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>TnpA expression for mating out, LS IS200 5’UTR</td>
</tr>
<tr>
<td>pDH897</td>
<td>pBAD24-tnp&lt;sub&gt;L&lt;/sub&gt;S; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>High-copy titrator, no cis-art200</td>
</tr>
<tr>
<td>pDH898</td>
<td>pDH900 with T7 P&lt;sub&gt;A&lt;/sub&gt;1-tnp&lt;sub&gt;45-298&lt;/sub&gt;; Tet&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Low-copy titrator, no cis-art200</td>
</tr>
<tr>
<td>pDH899</td>
<td>pDH900 with P&lt;sub&gt;Tet&lt;/sub&gt;-tnp&lt;sub&gt;45-298&lt;/sub&gt;; Tet&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Vector control</td>
</tr>
<tr>
<td>pDH902</td>
<td>pDH900 with IS200&lt;sub&gt;45-298&lt;/sub&gt; ; Tet&lt;sup&gt;R&lt;/sup&gt;</td>
<td>trans-art200</td>
</tr>
<tr>
<td>pDH912</td>
<td>pDH902 with M1 mutations; Tet&lt;sup&gt;R&lt;/sup&gt;</td>
<td>trans-art200&lt;sup&gt;M1&lt;/sup&gt;</td>
</tr>
<tr>
<td>pDH914</td>
<td>pDH899 with M1 mutations; Tet&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Low-copy titrator&lt;sup&gt;M1&lt;/sup&gt;</td>
</tr>
<tr>
<td>pDH916</td>
<td>pDH861 with M1 mutations; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>tnpA&lt;sup&gt;M1&lt;/sup&gt;-lacZ TLF</td>
</tr>
<tr>
<td>pDH918</td>
<td>pDH880 with M1 mutations; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>tnpA&lt;sup&gt;M1&lt;/sup&gt;-lacZ TLF, no cis-art200</td>
</tr>
</tbody>
</table>
4.4.2 RNA footprinting and toeprinting

*In vitro* transcription templates were generated by PCR using plasmids pDH861 (WT *tnpA*1-173 and art200), pDH862 (LS *tnpA*1-173 and art200), and pDH916 (M1 *tnpA*1-173 and art200) and primers oDH450 and oDH394 (*tnpA*1-173) or oDH500 and oDH501 (art200). RNAs were generated by *in vitro* transcription and 5’ labeled with [*γ^32P*]-ATP as previously described (Ellis et al, 2015b). Wild-type Hfq was purified by heat treatment and poly(A) affinity purification (Ross et al, 2010). RNase, Pb^{2+}, and hydroxyl radical footprinting were performed essentially as previously described (Ellis et al, 2015a; Ellis et al, 2015b; Ross et al, 2013). For footprinting reactions studying art200-*tnpA* pairing, each RNA was denatured at 95°C for 2 min and snap-cooled on ice for 3 min. Ambion 10X RNA Structure Buffer was added to a final concentration of 1X (10 mM Tris-HCl (pH 7), 100 mM KCl, 10 mM MgCl₂), and the RNAs were incubated separately at 37°C for 5 min to fold before mixing. For the control reactions in Figure 4.4B where the RNAs were folded together (FT, lanes 6, 11, 16, and 21) 5’[^32P]-labeled *tnpA* and art200 were mixed before the denaturing step.

Toeprinting was performed essentially as previously described (Ellis et al, 2015a). Briefly, unlabeled *tnpA* (2 pmol) was annealed to 5’[^32P]-labeled oDH394 before incubation with purified Hfq (0-8 pmol hexamer) or art200 (30 pmol) at 37°C. This was followed by addition of the 30S ribosomal subunit (3.6 pmol) and then initiator f_{mes}-tRNA (10 pmol; Sigma-Aldrich) for a final volume of 10 µL. Reverse transcription reactions were carried out at 37°C for 10 min with 200U of SuperScript II (Invitrogen).

Following ethanol precipitation, samples were resuspended in denaturing load dye (95% [v/v] formamide, 0.5X TBE, 3% [w/v] xylene cyanol) and resolved on a 10% polyacrylamide gel containing 7M urea. Dried gels were exposed to a phosphorimager storage screen, imaged with a Storm imager and quantitated with ImageQuant (GE Healthcare).
4.4.3 β-galactosidase assays

Cells were grown in LB supplemented (where necessary for plasmid selection) with ampicillin (100 µg/mL) and tetracycline (10 µg/mL). Saturated overnight cultures were used to seed subcultures (1:40 dilution), which were grown to mid-log phase (OD$_{600}$ = 0.4-0.6). The Miller assay was performed as previously described (Ross et al, 2010).

4.4.4 RNA extraction, primer extension

Cells were grown in LB supplemented (where appropriate for plasmid maintenance) with ampicillin (100 µg/mL) and/or tetracycline (10 µg/mL) to OD$_{600}$ = 0.6 at which time total RNA was extracted with acid phenol as previously described (Ellis et al, 2015a). 10 µg of total RNA was subject to primer extension analysis with 5’$^{32}$P-labeled oDH427 (Figure 4.2 and 4.3B) or oDH537 (Figure 4.6) (art200), oDH428 (tnpA), and oDH390 (lpp) and SuperScript III (Invitrogen) according to the manufacturer’s instructions. Following ethanol precipitation, samples were resuspended in denaturing load dye and resolved on a 10% polyacrylamide gel containing 7M urea. Dried gels were exposed to a phosphorimager storage screen, imaged with a Storm imager and quantitated with ImageQuant (GE Healthcare).

4.4.5 Conjugal mating out assay

The conjugal mating out assay was performed as previously described (Ellis et al, 2015a; Ross et al, 2010); see Figure S4.5 for schematic. Briefly, DBH291 was transformed with pDH857, pDH860, pDH896, or pDH897 and grown on LB agar plates containing ampicillin (100 µg/mL), kanamycin (25 µg/mL) and 0.05% arabinose (w/v). For the experiments presented in Figure 9C, DBH291 was transformed with pDH897 and pDH900 (vector) or pDH898 (titrator) and grown on LB agar plates containing ampicillin, tetracycline (10 µg/mL) and 0.05% arabinose (w/v). Individual colonies (‘donors’) were grown to saturation in LB containing (where appropriate for plasmid selection) ampicillin and tetracycline with 0.05% arabinose (w/v) and were subcultured 1:20 into LB containing 0.2% arabinose. Following mating with the recipient strain (DBH13), cells were plated on M9 glucose plates supplemented with thiamine, leucine
and streptomycin (150 µg/mL) (‘exconjugants’) or streptomycin and kanamycin (‘hops’). Transposition frequency was determined by dividing ‘hops’ by ‘exconjugants’.

4.5 Supplemental Material

4.5.1 Materials and Methods

4.5.1.1 Strain and plasmid construction

Supplemental plasmids and phage are listed in Table S4.1. The entire IS200 element (tnpA_6, STM4311) was amplified from Salmonella Typhimurium LT2 genomic DNA using GoTaq (Promega) and primers oDH378 and oDH379; the amplicon was cloned into pGem-T Easy (Promega) to produce pDH882 which served as the IS200 template for all further plasmid construction.

The WT IS200-lacZ translational fusion (TLF) was made by first amplifying IS200 from pDH882 with primers oDH130 and oDH535 and then cloning the amplicon into the Ncol/PstI sites of pGEM-T Easy; the lacZ gene was amplified with primers oDH534 and oDH536 and cloned into the PstI site to produce pDH861 which contains the first 323nt of IS200 fused in frame to codon 10 of lacZ. All mutations were introduced into pDH861 (or pDH880 for pDH918) using overlap PCR with primers oDH130 and oDH431 and the relevant mutagenic primers; pA1 (oDH414, oDH415), pA2 (oDH410, oDH411), pA3 (oDH467, oDH468), pA4 (oDH469, oDH470), pA5 (oDH471, oDH472), pA6 (oDH473, oDH474), LS (oDH416, oDH417) and M1 (oDH538, oDH539). PCR products were digested with EcoRI and cloned into the same sites in pDH861 resulting in pDH827, pDH863, pDH877, pDH878, pDH879, pDH880, pDH862, and pDH916, respectively.

To make the titrator plasmids, the SgrS terminator (amplified from DBH323 genomic DNA with oDH439 and oDH440) was cloned into the IS200 EcoRI site of a pDH880 partial digest. Next, a PCR amplicon containing bp 45-298 of IS200 and the sgrS terminator was made using Q5 polymerase (NEB) and primers oDH440 and either oDH531 (contains P_{tet}) or oDH530 (contains T7 P_{Al}). The M1 mutation was introduced
using overlap PCR with primers oDH440, oDH538, oDH539, and oDH531. PCR products were cloned into XmnI/ScaI digested pACYC184 resulting in plasmids pDH899 (low-copy titrator), pDH898 (high-copy titrator), and pDH914 (low-copy M1 titrator).

The trans-art200 plasmids were made by cloning the EcoRI/XmnI fragment of pDH882 (WT) or pDH916 (M1) into the same sites of pACYC184 to produce pDH902 and pDH914.

For mating out experiments, DBH33 was lysogenized with λDBH881 to create DBH291 (DBH33 Mini IS200-kan) (as described in (Ellis et al, 2015a; Ross et al, 2014)). First, pDH882 (pGEM-IS200) was digested with EcoRV, treated with Bal31 and ligated to kan^R to make pDH855. Mini IS200-kan was then cloned into Sall/SphI digested pACYC184 to make pDH856. Next, Mini IS200-kan was PCR amplified with primers oDH434 and oDH435 and the amplicon was digested with BamHI and cloned into BclI digested pNK81 to make pDH881. Finally, pDH881 was crossed onto λNK1039 (Haniford et al, 1989) to make λDBH881.

Transposase expression in mating out experiments was from pBAD24 derivatives (Guzman et al, 1995). pDH857 was made by amplifying the tnpA ORF from pDH855 with primers oDH451 and oDH452 and cloning the NcoI digested amplicon into the same site of pBAD24. The Y125F mutation was introduced into pDH857 using overlap PCR with primers oDH420, oDH421, oDH451, and oDH452 to make pDH860. Plasmids expressing TnpA under control of the WT (pDH896) or LS (pDH897) IS200 5'UTR were made by replacing the EcoRI/SphI fragment of pDH815 (WT) or pDH835 (LS) with the same fragment from pDH857. pDH815 and pDH835 consist of nts 45-323 of IS200 (WT or LS) fused in frame to codon 10 of lacZ and cloned into the NcoI/PstI sites of pBAD24. IS200 was amplified from pDH882 with primers oDH532 and oDH535 and cloned into the NcoI/PstI sites of pBAD24 and codon 10 of the lacZ gene (amplified with primers oDH534 and oDH536) was cloned into the PstI site of this plasmid. Finally, the pBAD24 Shine-Dalgarno sequence (SDBAD24) was mutated from AGGAGG to ACCACC to produce pDH815. The LS mutation was introduced to pDH815 by overlap PCR (primers oDH532, oDH431, oDH416, and oDH417) to produce pDH835.
4.5.1.2  ST-PCR

Semi-random, two-step PCR (ST-PCR) was used as previously described to map putative IS200 insertion sites (Chun et al, 1997; Ross et al, 2014). Briefly, genomic DNA from ‘hop’ colonies was amplified with a semi-random primer (oDH167) and a kan-specific primer (oDH388). PCR products were cloned into pGEM-T easy and sequenced, and then mapped to the E. coli MG1655 genome to identify insertion sites. Insertion site specific primers were then designed to flank the insertion site (oDH457, oDH458, oDH461, and oDH462) which was then amplified, cloned as above, and sequenced. The precise junction between each end of IS200 and the insertion site was then determined.

4.5.1.3  Southern blot

Genomic DNA was purified from ‘hop’ colonies (Sm\textsuperscript{R}Kan\textsuperscript{R}) using a Gen Elute Bacterial Genomic DNA kit (Sigma-Aldrich). DNA was digested with XmnI which cuts once in IS200 and then resolved on a 1% agarose gel. Southern blot for the kan\textsuperscript{R} gene was performed as previously described (Ross et al, 2014) with a \textsuperscript{32}P-labeled riboprobe.
4.6 Supplemental Figures

RNA isolated from isogenic \( hfq^+ / hfq^- \) strains was analyzed by primer extension. To facilitate detection of the \( tnpA-lacZ \) transcript the fusion was expressed from the \( P_{BAD} \) promoter. As \( tnpA \) levels did not increase in \( hfq^- \) versus \( hfq^+ \) we conclude that increased \( tnpA \) expression in \( hfq^- \) is not due to an increase in steady-state transcript levels.

Figure S 4.1. Impact of Hfq on \( tnpA-lacZ \) transcript levels.
Figure S 4.2. Lead and RNase structure probing of art200.
(A) RNase T1 and V1 structure probing of 5’32P-labeled art200 (90 nM). After folding at 37°C, art200 was mixed with dilute RNase T1 or V1 and incubated at ambient temperature for 15 min. Positions cleaved by RNase V1 or T1 are indicated with orange and purple arrows respectively. Positions sensitive to both RNases are indicated with grey arrows. Note that certain lanes from this gel were removed for clarity (indicated with white line) (B) Structure constraints from (A) were used with mFold (Zuker, 2003) to predict the secondary structure of art200. Coloured arrows show RNase sensitive positions. All nucleotide numbering is relative to the transcriptional start site for art200. (C) Lanes 1-3 of Figure 4.4A. 5’32P-art200 (69 nM) was allowed to fold at 37°C before limited cleavage with Pb2+ (lane 3). An RNase T1 sequencing reaction (G, lane 1) and untreated RNA (lane 2) are shown. Reactivity of each position was scored as strong (red), moderate (green), or weak (blue); scoring was relative to the most reactive position (C44) and background at each position (lane 2). Lead reactivity was modeled onto the secondary structure in (B, coloured letters).
Figure S 4.3. RNase structure probing of \textit{tnpA}\textsubscript{1-173}.

\textit{5'\textsuperscript{32}P}-labeled \textit{tnpA}\textsubscript{1-173} was subject to limited cleavage by RNase A, T1, or V1 for 15 min at ambient temperature. Samples were resolved on a 10\% denaturing polyacrylamide gel. An RNase T1 sequencing lane (lane 1) and untreated RNA (lane 2) are shown. Each position was assessed for reactivity to dsRNA specific RNase (V1; blue) or ssRNA specific RNase (A/T1; red). Positions sensitive to both types of RNase are highlighted in purple. Positions that were highly sensitive to only one type of RNase were used as hard
constraints in mFold (Zuker, 2003) to produce the secondary structure shown on the right. The reactivity of each nucleotide is shown on the secondary structure by coloured letters. All nucleotide numbering is relative to the *tnpA* translational start codon. The Shine-Dalgarno sequence and translational start codon are indicated in bold.
Figure S 4.4. Hfq blocks ribosome binding to tnpA in vitro.

The impact of Hfq on ribosome binding to tnpA1-173 was measured by a toeprinting assay (see also Figure 4.8A,B, 'experiment B'). The combined toeprint signal (G+15, A+16, and G+17) was normalized to positions +25 and +26 which served as loading controls. For example, lane 7 has 5-fold more signal at +25/+26 than the average signal for lanes 8, 9, 11, and 12. Note that these positions are 3' to the ribosome-tnpA interaction and therefore primer extension to this point should be insensitive to ribosome binding.
Figure S 4.5. Schematic of conjugal “mating out” assay for IS200 transposition.

DBH291 (‘donor strain’) contains the Mini IS200-kan element integrated in the chromosome as well as an F’ episome marked with lac-pro. DBH291 is transformed with a plasmid expressing TnpA under control of the PBAD promoter and various regulatory elements (constructs i-iv, see Figure 4.9). TnpA expressed from this plasmid can then catalyze transposition of Mini IS200-kan in trans. The donor strain is mated with DBH13 (recipient strain) which is strR and pro-. Following mating, cells are plated on M9 media containing streptomycin but not proline (to select for DBH13 cells which have acquired the F’, "exconjugates"), or streptomycin and kanamycin (to select for cells containing the F’ with IS200-kan, “hops”). The number of “exconjugates” indicates the mating efficiency and the ratio of “hops” to “exconjugates” is the relative transposition...
frequency. Note that this assay only measures transposition from the chromosome onto the F’ episome.
Figure S 4.6. Validation of IS200 transposition events (‘hops’).

(A) The location of the IS200-kan insertion site from two ‘hop’ colonies (construct i; Figure 4.9A) was determined by ST-PCR. The sequence of the two insertion sites (yafX and ybaA) is shown in black and the ‘left-end’ (LE) and ‘right-end’ (RE) sequence from IS200 is shown in red and blue, respectively. (B) IS200 transposition events (construct iii and iv; Figure 4.9A) were detected by Southern blot analysis (using a kan$^\text{R}$ probe) of genomic DNA that had been subject to restriction enzyme digestion. a-c refers to different mating mixes that gave rise to hop colonies. The fragment in lane 3 defines the starting location of IS200-kan in the chromosome of the donor strain. The appearance of fragments of different size in hop colony genomic preparations was taken as confirmation.
that transposition had occurred. Additionally, the observation that all 4 hop colonies from mating mix a produced a fragment of (apparently) the same size was taken as evidence that these are not independent transposition events. pMini contained genomic DNA from the donor strain harbouring IS200-kan on a plasmid (pDH881) and lane 2 contained genomic DNA isolated from the donor strain not containing IS200-kan. (C) Southern blot analysis of hop colonies from the indicated mating mixes (d-g) described in Figure 4.9C. a and b contain genomic DNA from hop colonies derived from mating mixes described in part B.
Figure S 4.7. Impact of mutations to a putative translational enhancer in *tnpA* on *tnpA* expression.

Mutations to C/A rich region (mutations in red, C/A rich region highlighted in blue) and the surrounding sequence were introduced into an otherwise wild-type *tnpA-lacZ* TLF (pDH861). β-galactosidase activity was measured by the Miller Assay from three independent clones and error bars show standard error on the mean.
Figure S 4.8. Predicted alternative secondary structures of the first 120nt of *tnpA*.

mFold was used to predict secondary structures of the 5’end of *tnpA* with a 50nt folding window and no other structural constraints (top left) or with the Shine-Dalgarno (SD)
sequence forced to be single-stranded. The folding energies of the unconstrained and top three alternative structures are shown. The Shine-Dalgarno (SD, blue) and complementary sequence (anti-SD, red) are highlighted in each structure along with the translational start codon (purple).
<table>
<thead>
<tr>
<th>Name</th>
<th>Relevant genotype</th>
<th>Use/ Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>λ Phage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>λNK1039</td>
<td>Encodes <em>his</em> operon</td>
<td>Plasmid-λ crosses (Haniford et al., 1989)</td>
</tr>
<tr>
<td>λDBH881</td>
<td>λNK1039 with Mini IS200-kan</td>
<td>Product of recombination between pDH881 and λNK1039; used for DBH291 construction</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pNK81</td>
<td>pBR333 derivative, encodes <em>his</em> operon; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>λ crosses</td>
</tr>
<tr>
<td>pDH815</td>
<td>pBAD24-<em>tnpA-lacZ</em> TLF; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Arabinose inducible TLF</td>
</tr>
<tr>
<td>pDH827</td>
<td>pDH815 with pA1 mutation; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Mutations to <em>P&lt;sub&gt;art200&lt;/sub&gt;</em> -10 region</td>
</tr>
<tr>
<td>pDH832</td>
<td>pDH815 with pA2 mutation; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Mutations to <em>P&lt;sub&gt;art200&lt;/sub&gt;</em> -10 region</td>
</tr>
<tr>
<td>pDH835</td>
<td>pDH815 with LS mutation; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Constructing pDH897</td>
</tr>
<tr>
<td>pDH855</td>
<td>pGEM-T easy with Mini IS200-kan; Kan&lt;sup&gt;R&lt;/sup&gt;Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Constructing DBH291</td>
</tr>
<tr>
<td>pDH856</td>
<td>pACYC184 with Mini IS200-kan; Kan&lt;sup&gt;R&lt;/sup&gt;Cm&lt;sup&gt;R&lt;/sup&gt;Tet&lt;sup&gt;S&lt;/sup&gt;</td>
<td>Constructing DBH291</td>
</tr>
<tr>
<td>pDH863</td>
<td>pDH861 with pA2 mutations; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Mutations to <em>P&lt;sub&gt;art200&lt;/sub&gt;</em> -10 region</td>
</tr>
<tr>
<td>pDH877</td>
<td>pDH861 with pA3 mutation; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Mutations to <em>P&lt;sub&gt;art200&lt;/sub&gt;</em> -10 region</td>
</tr>
<tr>
<td>pDH878</td>
<td>pDH861 with pA4 mutation; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Mutations to <em>P&lt;sub&gt;art200&lt;/sub&gt;</em> -10 region</td>
</tr>
<tr>
<td>pDH879</td>
<td>pDH861 with pA5 mutation; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Mutations to <em>P&lt;sub&gt;art200&lt;/sub&gt;</em> -35 region</td>
</tr>
<tr>
<td>pDH881</td>
<td>pNK81 with Mini IS200-kan; Kan&lt;sup&gt;R&lt;/sup&gt;Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Crossing Mini IS200-kan onto λNK1039</td>
</tr>
<tr>
<td>pDH882</td>
<td>IS200 cloned into pGem-T easy; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Source of IS200 sequence</td>
</tr>
</tbody>
</table>
### Table S 4.2. List of oligonucleotides for Chapter 4.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’ to 3’)</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>oDH130</td>
<td>GAATTGTAATACGACTCACTATA TAG</td>
<td>TLF construction</td>
</tr>
<tr>
<td>oDH167</td>
<td>GGCACCGCTGACTAGTACNANNNNNNNNNGAT AT</td>
<td>ST-PCR - random primer</td>
</tr>
<tr>
<td>oDH378</td>
<td>TTGTGGTGTTAGCCTTAACGGTAAAT T G</td>
<td>IS200 amplification from LT2</td>
</tr>
<tr>
<td>oDH379</td>
<td>GGCATTCGGTAACTGGTATGGA AG</td>
<td></td>
</tr>
<tr>
<td>oDH388</td>
<td>AGGCATATAGGCCAGTCTTAACCAAATTAACCATTTCTA G</td>
<td>ST-PCR - Kan specific</td>
</tr>
<tr>
<td>oDH390</td>
<td>CAGCTGGTCAACTTTAGCAGTTGACAG</td>
<td></td>
</tr>
<tr>
<td>oDH394</td>
<td>TATTTGGGCGCGAAACTATG</td>
<td></td>
</tr>
<tr>
<td>oDH397</td>
<td>GCTGCTACTGCCTACCGTTCTC</td>
<td>Sequencing IS200 plasmids</td>
</tr>
<tr>
<td>oDH410</td>
<td>GGGGACGAGGGGGGCTTTGGG</td>
<td>pDH863 construction</td>
</tr>
<tr>
<td>oDH411</td>
<td>CGCCAAGCCTCCTGTCCTCC</td>
<td></td>
</tr>
<tr>
<td>oDH414</td>
<td>GGGGACGAGCCCGACCTTAGG</td>
<td>pDH827 construction</td>
</tr>
<tr>
<td>oDH415</td>
<td>CGCTAAGCTCGGTGCTCC</td>
<td></td>
</tr>
<tr>
<td>oDH416</td>
<td>CCAGTTATTAAAGGGAATGATTTGGTTAAAA C</td>
<td>pDH862 construction</td>
</tr>
<tr>
<td>oDH417</td>
<td>GTTTTAACAAATCATTTCCTTTTAAATAACTG G</td>
<td></td>
</tr>
<tr>
<td>oDH420</td>
<td>GATACAGGACTTCTCATAAAGC</td>
<td>pDH860 construction</td>
</tr>
<tr>
<td>oDH421</td>
<td>GTGCTTTATGAATGCTGTATC</td>
<td>pDH860 construction</td>
</tr>
<tr>
<td>oDH427</td>
<td>CATCTTGGGCTCTGGGAC</td>
<td>art200 primer extension</td>
</tr>
<tr>
<td>oDH428</td>
<td>GTTGCCAGAAGCGCAAGT</td>
<td>tnpA primer extension</td>
</tr>
<tr>
<td>oDH431</td>
<td>AGGCTGCGAAGCTGTGGG</td>
<td>TLF construction</td>
</tr>
<tr>
<td>oDH434</td>
<td>GCGGGATCCGCCGCCGGGTGTTTTGTTCTATG</td>
<td>BamHI IS200-LE</td>
</tr>
<tr>
<td>oDH435</td>
<td>GAGGGATCCAAATTTATAAAATATATATCCTC</td>
<td>BamHI IS200-RE</td>
</tr>
<tr>
<td>oDH439</td>
<td>TGCTCGGAATTTCTGAAAGTGGAC T T</td>
<td>titrator construction</td>
</tr>
<tr>
<td>oDH440</td>
<td>TAGGCCAGAATTTCTAACGAGTAAGG</td>
<td></td>
</tr>
<tr>
<td>oDH450</td>
<td>TAATACTGACTCATACTAGGCTTTCAGCTTTAAG CCAG</td>
<td></td>
</tr>
<tr>
<td>oDH451</td>
<td>ATACCATGGGGGAGCGAAAGAGCAGGCTAGC</td>
<td></td>
</tr>
<tr>
<td>oDH452</td>
<td>ATACCATGGGGACCTTGATATGTAACGGC</td>
<td></td>
</tr>
<tr>
<td>oDH453</td>
<td>ACCGTATTACACGTGTAATTCG G</td>
<td>Southern Probe template (Kan rev)</td>
</tr>
<tr>
<td>oDH454</td>
<td>TAAATACCAGCTCAGCTATAGGGAGCGTTGCAAT ATGTTACA</td>
<td>Southern Probe template (T7-Kan)</td>
</tr>
<tr>
<td>oDH457</td>
<td>ATACATTTAACATGAGAAGACGAA</td>
<td>IS200 Hop verification (ybaA)</td>
</tr>
<tr>
<td>oDH458</td>
<td>TCAGCGAAGAGTTAATAATTTTCG</td>
<td>IS200 Hop verification (ybaA)</td>
</tr>
<tr>
<td>oDH461</td>
<td>TGCGTCTGTGTGTTGTCAAC</td>
<td>IS200 Hop verification (yafX)</td>
</tr>
<tr>
<td>oDH462</td>
<td>AGGCATCCGTACAGACACGG</td>
<td>IS200 Hop verification (yafX)</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>oDH467</td>
<td>GGGGGACGAACGCAGGCTTAGCG</td>
<td>pDH877 construction</td>
</tr>
<tr>
<td>oDH468</td>
<td>CGCTAAGCCGGCTTGCTCCCCC</td>
<td>pDH878 construction</td>
</tr>
<tr>
<td>oDH469</td>
<td>GGGGAGGCAGCGGCTTAGCG</td>
<td>pDH879 construction</td>
</tr>
<tr>
<td>oDH470</td>
<td>CGCTAAGCCGGCCGCTCCCCC</td>
<td>pDH880 construction</td>
</tr>
<tr>
<td>oDH471</td>
<td>GCAACCCCGAGCAGGCTTAATATG</td>
<td>Fwd primer for art200 in vitro transcription template, contains T7 promoter sequence</td>
</tr>
<tr>
<td>oDH472</td>
<td>GATAATTTACACCCGCTCGGTGTC</td>
<td>Rev primer for art200 in vitro transcription template</td>
</tr>
<tr>
<td>oDH473</td>
<td>CGCATGGGACCCGCCTATACACATG</td>
<td>pDH898 construction</td>
</tr>
<tr>
<td>oDH474</td>
<td>CTATGTGATAGCCCGGTTCATCCCG</td>
<td>pDH899 construction</td>
</tr>
<tr>
<td>oDH500</td>
<td>TAATACGACTATAGGTCCCCCATGGGACC</td>
<td>Fwd primer for art200 in vitro transcription template</td>
</tr>
<tr>
<td>oDH501</td>
<td>AACCCCTTTTGTATTGTTAAAAC</td>
<td>Rev primer for art200 in vitro transcription template</td>
</tr>
<tr>
<td>oDH530</td>
<td>TTATCAAAAAAGGTATTTGACCTAAAGTCTAACC TATAGGATACCTTCAGCCTCTCAGCTTTAAGCCAG TTATTTAA</td>
<td>pDH861 construction</td>
</tr>
<tr>
<td>oDH531</td>
<td>AAGAATTTCATGTTGACAGCTTTATCATCATGAT AAGCTTTTAATGCGGTAGTCAGCTTTAAGCCAGTT TATTTAA</td>
<td>art200 primer extension</td>
</tr>
<tr>
<td>oDH532</td>
<td>CACCATGGTCAAGCTTTAAGCCAGTTA</td>
<td>Mutagenic primers for “M1” mutation</td>
</tr>
<tr>
<td>oDH534</td>
<td>GCCTGCAGGCTTTAACAACGTCTGAC</td>
<td></td>
</tr>
<tr>
<td>oDH535</td>
<td>GTCTGCAATCTGCAACAAATCTCTGC</td>
<td></td>
</tr>
<tr>
<td>oDH536</td>
<td>GCCTGCAGAGATTATTTTGACACAGA</td>
<td></td>
</tr>
<tr>
<td>oDH537</td>
<td>TGATTTGGTTAACAACATCTTGGCG</td>
<td></td>
</tr>
<tr>
<td>oDH538</td>
<td>GTTTAACAATCTTGGCTTCAGCTCCGAACTGCAAAAA GTTCAAC</td>
<td></td>
</tr>
<tr>
<td>oDH539</td>
<td>GTTGAACCTTTTGCAGTTCGGAGACCGCAAGATG TTTTAAC</td>
<td></td>
</tr>
</tbody>
</table>
4.7 References


Ellis MJ, Trussler RS, Haniford DB (2015a) Hfq binds directly to the ribosome-binding site of IS10 transposase mRNA to inhibit translation. Molecular microbiology 96: 633-650


Haniford DB, Chelouche AR, Kleckner N (1989) A specific class of IS10 transposase mutants are blocked for target site interactions and promote formation of an excised transposon fragment. Cell 59: 385-394


Kittle JD, Simons RW, Lee J, Kleckner N (1989) Insertion sequence IS10 anti-sense pairing initiates by an interaction between the 5' end of the target RNA and a loop in the anti-sense RNA. *Journal of molecular biology* **210**: 561-572


Ross JA, Trussler RS, Black MD, McLellan CR, Haniford DB (2014) Tn5 transposition in Escherichia coli is repressed by Hfq and activated by over-expression of the small non-coding RNA SgrS. *Mob DNA* **5**: 27


Siguer P, Gourbeyre E, Chandler M (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS microbiology reviews* 38: 865-891


Chapter 5

5 A transposon derived small RNA regulates gene expression in *Salmonella* Typhimurium

5.1 Introduction

IS200 is the smallest prokaryotic transposon and is widely conserved in Enterobacteriaceae and found throughout Eubacteria and Archaea. One unusual feature of IS200 elements is the high copy-number achieved in *Yersinia* and *Salmonella* spp. (Beuzon et al, 2004; Ellis et al, 2015a; Siguier et al, 2006). Many strains of *Y. pestis* contain more than 50 copies of the IS200 ortholog IS1541, while strains of *Salmonella enterica* subsp. *enterica* serovar Typhimurium (*S. Typhimurium*) typically contain 5-12 copies and *S. Typhi* contains 26 copies of IS200 per genome. In the above cases, all IS200 paralogs are 100% conserved and in general IS200 orthologs share >90% identity. A highly active transposon might be expected to achieve this high copy-number and repeated transposition would maintain sequence identity of paralogs; however IS200 is an essentially dormant transposon (Beuzon et al, 2004; Ellis et al, 2015a). Conservation and copy-number might therefore reflect a selective pressure on the host bacterium to maintain IS200. Transposons can contribute to host fitness in several ways including: (i) by mediating DNA rearrangements that influence host gene expression and gene structure (Siguier et al, 2014); (ii) contributing passenger genes such as antibiotic resistance determinants (Davies, 1994); (iii) providing a rich source of DNA regulatory sequence (Feschotte, 2008); (iv) providing proteins and/or protein motifs from transposase proteins that can be domesticated by the host (Volff, 2006); and (v) providing regulatory RNAs that affect host gene expression (Gomes-Filho et al, 2015; Martens et al, 2013). As a simple insertion sequence, IS200 does not encode any passenger genes, and the dormancy

---

4 The work presented in this chapter has been submitted for publication and is currently in revision: Ellis MJ, Trussler RS, Haniford DB. A transposon derived small RNA regulates gene expression in *Salmonella* Typhimurium. *Nucleic Acids Research*. NAR-02816-Y-2016
of IS200 suggests that this element would not contribute transposition-dependent functions to the host.

Non-coding RNAs (ncRNA) play a crucial role in regulating many critical processes in bacteria including outer membrane homeostasis, metabolism, and virulence (Storz et al, 2011; Wagner & Romby, 2015). The largest class of bacterial ncRNA are small RNAs (sRNA) that base-pair with target mRNAs and affect translation and/or transcript stability. sRNAs are typically expressed from intergenic regions and therefore have limited sequence complementarity with their trans-encoded targets. A related class of ncRNA are antisense RNAs (asRNA) which are encoded on the opposite strand of DNA to their target mRNA. Accordingly, asRNAs have much more extensive complementarity with their cis-encoded targets. The third and smallest class of bacterial ncRNAs act by binding to and regulating protein activity (e.g. 6S RNA, CsrB/C). The classic distinction between these three classes of ncRNA has been challenged with continually emerging examples of dual-function ncRNA, including sRNAs derived from mRNAs (Chao et al, 2012; Chao & Vogel, 2016; Guo et al, 2014; Hershko-Shalev et al, 2016), base-pairing sRNAs acting to modulate protein activity (Ellis et al, 2015b; Holmqvist et al, 2016; Jørgensen et al, 2013; van Nues et al, 2016), and asRNAs acting in trans to regulate genes expressed from different loci (Jager et al, 2012; Sayed et al, 2012).

One common feature for base-pairing ncRNAs is that the RNA-binding protein Hfq is typically required to facilitate pairing when there is limited complementarity between an sRNA and mRNA (Vogel & Luisi, 2011). In general, an interaction between Hfq and an mRNA or sRNA indicates that the RNA is involved in post-transcriptional regulation via a base-pairing mechanism.

IS200 elements express two RNA molecules (Figure 5.1A), the first is an mRNA encoding the transposase protein (*tnpA*), and the second is an asRNA (art200, previously named STnc490) that is complementary to the *tnpA* 5’UTR (Beuzon et al, 1999; Ellis et al, 2015a; Sittka et al, 2008). Expression of the IS200 TnpA is strongly repressed by 4 independent mechanisms. First, the left-end of IS200 contains an inverted repeat that forms a strong, bi-directional, Rho-independent transcriptional terminator. This regulatory element ensures that impinging transcription does not activate TnpA.
Figure 5.1. IS200 and experimental approach.

(A) IS200 encodes a transposase mRNA (tnpA, red) and an antisense RNA (art200, blue). The tnpA\textsubscript{trunc}-255 transcript encodes the first 255 nt of tnpA fused to the last 108nt of SgrS (black, includes an intrinsic terminator) and is expressed from the Tet promoter. (B) Approach used to deplete art200. Pairing between tnpA (red) and art200 (blue) results in degradation of art200. The M1 mutation alters three critical nucleotides in the terminal loop of tnpA and prevents pairing with art200. (C) Heat map showing expected expression of IS200 RNAs in S. Typhimurium LT2 containing plasmids over-expressing either WT or M1 forms of tnpA\textsubscript{trunc}-255. Note that ‘tnpA’ signifies endogenous transposase transcript and ‘vector’ is a control plasmid that does not overexpress tnpA.
expression and terminates ~85% of upstream transcripts (Beuzon et al, 1999). Second, translation of *tnpA* is strongly repressed by mRNA secondary structure that includes the Shine-Dalgarno sequence (SD). This stem-loop element represses *tnpA* expression 20-fold by preventing 30S ribosome binding. Third, art200 base-pairs with *tnpA* to inhibit ribosome binding, and reduces translation 15-fold. Lastly, *tnpA* translation is inhibited directly by the RNA-binding protein Hfq, which recognizes a sequence immediately upstream of the SD and accordingly stericly occludes ribosome binding. The three post-transcriptional mechanisms act independently and together suppress translation of *tnpA* by at least 750-fold, ensuring almost no TnpA protein is produced (Ellis et al, 2015a).

While these regulatory mechanisms appear to be redundant, *tnpA* expression is reasonably high in *S. Typhimurium* for a transposon (~10% the expression of *hfq* in mid-exponential phase (Kroger et al, 2013)). It therefore appears that IS200 elements have evolved to maintain moderate transcription of *tnpA* from an IS200 encoded promoter, but close to no synthesis of TnpA. Another noteworthy feature of IS200-encoded RNAs is that art200 expression appears to be growth phase regulated, with increased expression when *S. Typhimurium* transitions to stationary phase in rich media, as well as in growth media that stimulate *Salmonella* pathogenicity island (SPI) expression (Figure S5.1; (Sittka et al, 2008)). Additionally, art200 interacts with Hfq *in vivo*, although Hfq is dispensable for antisense regulation of *tnpA* expression. Intriguingly, while art200 expression is increased in stationary-phase, *tnpA* expression decreases ~5-fold (Kroger et al, 2013), which may indicate that art200 expression is altered to control *tnpA* RNA levels. One explanation for the unusual characteristics of IS200-encoded RNAs is that a moderately expressed but never translated *tnpA* provides a way in which IS200 transposition could be rapidly activated under certain conditions. However, previous work found that IS200 transposition is remarkably rare, even when post-transcriptional regulation is completely eliminated (Ellis et al, 2015a). With respect to art200’s expression patterns and Hfq-binding properties, this could simply reflect stochastic evolution of the promoter and sequence of a regulatory RNA. A more intriguing explanation for the peculiar properties of *tnpA* and art200 is that one or both IS200-encoded RNAs serves a regulatory role independent of controlling transposition. In this
scenario, an IS200 encoded RNA might provide a selective advantage to *Salmonella* spp. and accordingly explain the conservation and high copy-number of this transposon.

In the current work we performed an RNA-Seq experiment to ask if IS200-encoded RNAs affect gene expression in *S. Typhimurium*. We provide evidence that the 5’UTR of *tnpA* represses many genes including the SPI-1 encoded transcription factor, *invF*. Our data suggests that *tnpA* base-pairs with *invF*, and the consequence of this interaction is down-regulation of the SPI-1 translocon (*sicAsipBC*). This work is the first demonstration of a bacterial transposon encoding regulatory RNAs that influence host gene expression.

### 5.2 Results

**5.2.1 Profiling changes in *S. Typhimurium* gene expression in response to altered levels of IS200-encoded transcripts**

We used RNA-Seq to analyze gene expression in *S. Typhimurium* LT2 under conditions where levels of *tnpA* and art200 were altered from native levels. In one strain we introduced a plasmid that constitutively over-expresses a truncated form (nt 1-255) of the transposase mRNA (*tnpA*<sub>trunc</sub><sup>WT</sup>-255). This strain produces very low amounts of art200 because *tnpA*<sub>trunc</sub><sup>WT</sup>-255 RNA pairs with art200 and this pairing promotes degradation of art200 (Figure 5.1B,C; Figure S5.1C). When we looked for differentially expressed genes in this strain vs an empty vector control strain, we identified 187 genes with altered expression (Figure 5.2A, black dots), 99 of which had at least a 2-fold change in expression. This altered pattern of gene expression could arise from either depletion of art200 and/or the overexpression of the truncated *tnpA* mRNA. To distinguish between these possibilities, we profiled gene expression in a third strain expressing a truncated form of *tnpA* (*tnpA*<sub>trunc</sub><sup>M1</sup>-255) that is unable to pair with art200 (Figure 5.2B). Genes affected by depletion of art200 would show differential expression when *tnpA*<sub>trunc</sub><sup>WT</sup>-255 was overexpressed but not when *tnpA*<sub>trunc</sub><sup>M1</sup>-255 was overexpressed. When all three comparisons were made, only 6 genes appeared to be uniquely regulated by art200 (Figure 5.2C; *glnH*, *gltI*, *acs*, *icdA*, *hutU*, and a predicted asRNA to the 3’end of *fadR*). In
Figure 5.2. Summary of RNA-Seq data.

(A,B) Expression plot comparing relative abundance (log$_2$ clr) of S. Typhimurium LT2 transcripts in the presence of an empty-vector (x-axis) or plasmid expressing WT (y-axis, A) or M1 (y-axis, B) $tnpA_{\text{trunc}}$-255. Differentially expressed genes (effect size $>$2) are indicated in black, and dotted lines indicate a two-fold change in expression from the line of best fit for the data (A, Pearson’s $r$ = 0.9393; B, Pearson’s $r$ = 0.9348). Reads derived from $tnpA_{\text{trunc}}$-255 mapped to either the IS200 transposase coding sequence ($tnpA$) or 5’UTR ($tnpA$ 5’UTR) and are indicated in blue. SPI-1 genes sicA, sipB, sipC, and invF are highlighted in red; note that invF was repressed >3-fold by $tnpA_{\text{trunc}}^{\text{WT}}$-255 but fell below out cut-off for differential expression (effect size = -1.2025). Genes with an effect size $<$ 2 are indicated in grey and are not considered to be differentially expressed. (C) Venn diagram showing the overlap of genes identified as differentially expressed when comparing the empty vector to $tnpA_{\text{trunc}}^{\text{WT}}$-255 (blue) or $tnpA_{\text{trunc}}^{\text{M1}}$-255 (yellow, or $tnpA_{\text{trunc}}^{\text{WT}}$-255 to $tnpA_{\text{trunc}}^{\text{M1}}$-255 (green). (D) Results of GO Enrichment Analysis. The log$_2$ fold change (Vector vs $tnpA_{\text{trunc}}^{\text{WT}}$-255) of genes in the three enriched biological processes are shown along with the enrichment score and p-value from the PANTHER Overrepresentation test. Horizontal bars indicate the median fold-change for each biological process.
contrast, genes regulated by $tnpA_{trunc}$ overexpression would show differential expression in both WT and M1 $tnpA_{trunc}$ strains when compared to an empty vector. A total of 73 genes fit this criteria (Figure 5.2C). Based on this analysis we concluded that transcripts derived from IS200 impact on host gene expression and that high levels of a truncated form of $tnpA$ that includes the 5’UTR has a greater impact on host gene expression than depletion of art200.

Lastly, we searched for cellular processes enriched with genes affected by $tnpA_{trunc}$ overexpression. This analysis found that $tnpA$ overexpression significantly represses genes involved in pathogenesis, glycerol-3-phosphate metabolism, and oxidation-reduction reactions (Figure 5.2D). The strongest change in gene expression in any of these pathways was the SPI-1 encoded effector protein, sipC (10-fold repression by overexpression of $tnpA_{trunc}^{WT}$-255). As S. Typhimurium LT2 is avirulent (Swords et al, 1997), we switched to the virulent SL1344 strain (7 copies of IS200) for subsequent studies.

5.2.2 Characterization of $tnpA$ derived RNAs

Our RNA-Seq analysis revealed that overexpression of the first 1/3 of transposase mRNA had a substantial impact on gene expression in S. Typhimurium. While this points to $tnpA$ mRNA acting as a regulatory RNA, we thought it more likely that a naturally truncated or processed form of $tnpA$ is produced from the 5’end to act as a regulatory RNA. This would be in line with other recently discovered mRNA derived sRNAs (Miyakoshi et al, 2015b). We initially looked for evidence of an sRNA derived from the 5’end by performing a Northern blot (5’UTR probe) on RNA isolated from a strain expressing native levels of $tnpA$ (WT) or a strain where $tnpA$ was over-expressed through the fusion of the pTet promoter to one copy of $tnpA$ in the chromosome. In the latter strain we detected three species, two of which are approximately 90 and 110 nts and the other is >310 nts (Figure 5.3A, lane 3). The 90 and >310 nt species were also just detectable in the strain expressing $tnpA$ at native levels (lane 1). In contrast, none of these species were detected in a strain where 4 of 7 copies of the $tnpA$ gene were deleted (lane 2).
Figure 5.3. Processing of the \textit{tnpA} transcript.

(A) A northern blot of \textit{tnpA} RNA isolated from SL1344 strains expressing \textit{tnpA} at endogenous levels (WT), over-expressing \textit{tnpA} from the \textit{tnpA}_7 locus (\textit{tnpA}_7::\textit{kan-}\textit{pTet}, \textit{pTet}), or with a reduced number of endogenous copies of \textit{tnpA} (\textit{\Delta}tnpA\textsubscript{2/4/6/7}, \textit{\Delta}tnpA). Full-length (closed circle) and processed (open circle) forms of \textit{tnpA} were detected with a probe that anneals to the 5’UTR (oDH429). 5S rRNA was used as a loading control.

(B,D) 5’ends of \textit{tnpA} were mapped using primer extension. RNA was isolated from the above strains (two replicates) and \textit{tnpA} was detected using a primer that anneals to the
5′UTR (nts 46-64, B) or coding sequence (nts 151-171, D). ddNTP sequencing lanes were used to determine the nucleotide position of primer extension products relative to the transcription start site (+1, ‘Full-length’). (C) *tnpA* is processed at U_{17} and A_{19}. RNA isolated from the WT or *tnpA_7::kan-pTet* strains was treated with TEX (+) or incubated with buffer (-) before *tnpA* was detected by primer extension. (E) Summary of primer extension experiments. The major primer extension products from part B and D are illustrated along with the primer binding sites. The two primers used for primer extension would detect different molecules of *tnpA* based on processing occurring between the primer binding sites. From the positions of 5′ends and the size of low molecular weight RNA species in the Northern (part A), we infer that the *tnpA* transcript is processed at two sites (red) to produce two stable 5′UTR-containing species (site B, *tnpA*-110; sites A+B, *tnpA*-90).
Additionally, both the 110 and 90nt species were detected by Northern blots on samples where \( tnpA_{\text{trunc-255}} \) was overexpressed (Figure S5.1C). Taken together these results show that: (i) the native \( tnpA \) gene generated one or more sRNAs, (ii) sRNA production does not require more than 255 bp of the \( tnpA \) gene, and (iii) sRNA production occurs independent of the promoter used to drive \( tnpA \) transcription. The latter point is suggestive of sRNAs being produced through RNA processing of the \( tnpA \) transcript.

We next performed primer extension on the above RNA samples to map 5’ends of each species. In one experiment we used a primer that anneals to the 5’UTR (nts 46-64). The results show that the majority of \( tnpA \) transcripts start at position 19 rather than the expected transcription start site (Figure 5.3B). This pattern was observed both when \( tnpA \) was over-expressed and expressed at native levels. We also show that prior treatment of the RNA with 5’monophosphate dependent terminator exonuclease (TEX) resulted in loss of the primer extension signal at nt 19, indicating that this 5’end is generated through transcript processing (Figure 5.3C). In a second experiment we used a primer that anneals in the coding sequence (nts 151-171) (Figure 5.3D). Here we also identified the position 19 5’end and additional 5’ends surrounding position 108. These alternative 5’ends were also lost upon TEX treatment, indicating processing in a second region of the \( tnpA \) transcript (Figure S5.2A). Processing events at positions 19 and 108 would generate a 5’UTR containing species of ~90 nt. In contrast, processing at only the downstream site would generate a 5’UTR containing species of ~110 nt in length. Based on these experiments, we infer that processing at sites designated A and B in Figure 5.3D generate stable \( tnpA \) encoded sRNAs (Figure 5.3E; Figure S5.2).

To test the hypothesis that one or both of the above described sRNAs are actually the active molecules for regulating host genes, we made additional \( tnpA_{\text{trunc}} \) constructs (first 50, 200, and 250 nt of \( tnpA \) overexpressed from plasmid) to determine the minimal \( tnpA \) required for affecting gene expression in \( S. \) Typhimurium; both \( tnpA_{\text{trunc-200}} \) and -250 are processed to produce ~110 and ~90 nt species (Figures S5.2D, S5.3A). We used RT-qPCR to determine which of these truncated \( tnpA \) molecules down-regulates a set of functionally related genes (\( sicA, sipB \) and \( sipC \)) identified in our RNA-Seq experiment to be repressed by \( tnpA \). All three truncated forms of \( tnpA \) down-regulated \( sicA, sipB \) and
sipC expression (> 2.5-fold) but not the expression of thrS, a gene whose expression was not affected by tnpA in the RNA-Seq analysis (Figure 5.4A). From this experiment it is evident that over-expression of only the first 50 nt of tnpA is sufficient to negatively regulate expression of the aforementioned genes, indicating that either the 110- or 90nt processed species is a functional sRNA. It may also be significant that of the three truncated forms of tnpA tested in this experiment, tnpA_trunc-50 down-regulated expression of the target genes to the highest degree and is the only one of the three tnpA RNAs incapable of base-pairing with art200 (Figure S5.2D). The latter point may be particularly relevant if art200 factors into tnpA transcript processing.

In the RT-qPCR analysis we also monitored tnpA expression using the primers shown in Figure S5.2C. These primers do not amplify tnpA species processed within the 5’UTR or the tnpA_trunc-50 RNA. However, they do provide a measure of the relative abundance of unprocessed tnpA produced in strains with a vector or overexpressing tnpA_trunc-200 or -250; in the latter two strains, tnpA was expressed 95- and 6.5-fold higher than endogenous tnpA respectively. This is particularly significant because both forms of tnpA down-regulated the target genes to the same extent and thus we concluded that tnpA levels do not need to be increased by more than 6.5-fold to see an impact on target gene expression.

sicA, sipB, and sipC are the first three genes in a large polycistronic transcript encoding secreted effector proteins for the SPI-1 type-III secretion system (T3SS) and are required for invasion of non-phagocytic cells (Lostroh & Lee, 2001). To gain insight into how tnpA regulates sicAsipBC we searched for predicted base-pairing interactions between this transcript and the 5’UTR of tnpA using TargetRNA2 (Kery et al, 2014) and IntaRNA (Wright et al, 2014) but no predicted interactions were found. Transcription of the sic/sip operon is activated directly by the SPI-1 encoded transcription factor InvF and the effect of tnpA on sicAsipBC could therefore be mediated through direct regulation of invF. Indeed, all three tnpA_trunc constructs repressed invF, with over-expression of tnpA_trunc-50 reducing invF mRNA levels 3.5-fold (Figure 5.4A). We also examined the effect of constitutive overexpression of tnpA_trunc on InvF protein levels with a strain of SL1344 containing a 3xFLAG tag integrated at the C-terminus of the native invF gene.
Figure 5.4. The 5’end of tnpA represses invF mRNA and protein expression.

(A) RT-qPCR was performed on S. Typhimurium SL1344 cells expressing different truncated forms of tnpA grown to late-exponential phase (OD$_{600} = 1.2$). Note that tnpA expression for samples containing pTet-tnpA$_{trunc}$-50 reflects the endogenous amount of tnpA as this construct lacks both primer binding sites for amplification of tnpA. (B) InvF Western blot on SL1344 cells expressing tnpA$_{trunc}$ constructs. Cells contained a 3xFLAG tag integrated at the end of the invF gene and cell extracts (prepared at mid-exponential phase; OD$_{600} = 0.8$) were probed with an α-FLAG antibody. Wild-type SL1344 cells provided a negative control, and GroES was used as a loading control. (C) RT-qPCR was performed on SL1344 strains expressing full-length tnpA from the native (WT) or Tet (tnpA$_7$::kan-pTet) promoter grown to late-exponential phase (OD$_{600} = 1.2$). (D) InvF Western blot on SL1344 strains expressing full length tnpA. Extracts were prepared at an OD$_{600}$ of 0.4 or 1.0 from cells expressing tnpA at endogenous levels (−) or constitutively overexpressed from the tnpA$_7$::cm-pTet locus (+). InvF was detected as in (B), and DnaK served as a loading control.

In all panels, error bars show the standard error on the mean for four (A,C) or three (B,D) biological replicates.
Consistent with our RT-qPCR analysis, \( tnpA_{\text{trunc}} \) repressed InvF protein levels over 2-fold (Figure 5.4B).

We also looked at the ability of \( tnpA \) to inhibit \( invF \) expression using over-expressed full-length \( tnpA \) (\( tnpA_{7::\text{kan-pTet}} \), Figure 5.4C,D). We show that in late exponential phase this strain expressed \( tnpA \) at a level ~65-fold higher than the WT strain, and decreased \( invF \) transcript and protein levels 2-2.5-fold. For comparison, \( invF \) levels were decreased 5.7-fold in a \( \Delta hilA \) strain. As HilA is a transcriptional activator of \( invF \), the \( \Delta hilA \) strain provides a measure of uninduced \( invF \) expression. Together, the above data indicates that a \( tnpA \)-derived sRNA inhibits expression of SPI-1 effector proteins \( sicA, sipB, sipC \) by repressing \( invF \) expression.

### 5.2.3 Over-expression of \( tnpA \) represses expression of SPI-1 in a growth phase dependent manner

We next asked if the regulation of SPI-1 genes by a \( tnpA \)-derived sRNA is linked to growth phase, as \( invF \) expression is induced in late exponential and early stationary phase.

We profiled the expression of \( invF \) and other SPI-1 encoded genes (\( sicA, sipB, sipC, \) and \( prgH \)) during 5 different growth phases in a WT or \( tnpA \) overexpression strain (\( tnpA_{7::\text{kan-pTet}} \)). Importantly, there was no difference in growth rate between the two strains (Figure S5.4A). Over-expression of \( tnpA \) did not affect SPI-1 gene expression in cells in lag- or early-exponential phase (Figure 5.5A,B). In both of these growth phases, \( tnpA \) in the WT strain was expressed higher than \( invF \) (Figure 5.5F), suggesting that the native expression of \( tnpA \) was sufficient for regulating \( invF \). Once cells reached late-exponential phase, overexpression of \( tnpA \) repressed \( invF \) (2-fold), \( sicA \) (5.5-fold), \( sipB \) (4-fold), and \( sipC \) (2-fold); \( prgH \) expression (an InvF-independent SPI-1 encoded gene) was not affected by \( tnpA \) overexpression (Figure 5.5C). At this growth phase \( invF \) is moderately induced (~6-fold) relative to early-exponential phase, and is now present at
Figure 5.5. Overexpression of tnpA RNA downregulates invF and sicAsipBC in a growth phase dependent manner.

RT-qPCR was performed on SL1344 cells (WT or tnpA_7::kan-pTet) grown to different growth phases. LB was inoculated with single colonies of the indicated strains and RNA was harvested after 18 hrs (E, deep-stationary phase). The 18 hr cultures were used to seed subcultures, and RNA was isolated after 1.25 hrs (A, lag-phase), 2 hrs (B, early-exponential phase), 3 hrs (C, late-exponential phase), or 4 hrs (D, early-stationary phase). Error bars show the standard error on the mean (n = 4). The relative amount of tnpA to invF (ΔΔCT) for WT or tnpA_7::kan-pTet strains is shown in a heat map (F). Raw ΔCT values (relative to 16S rRNA) for all genes and growth phases are shown in Figure S5.4.
~2-fold excess to \textit{tnpA} in the WT strain (Figure 5.5F). Here, endogenous \textit{tnpA} is presumably limiting, explaining why this growth phase shows the largest impact of \textit{tnpA} overexpression. Lastly, \textit{tnpA} overexpression had only a minor effect on \textit{invF} expression during early- and deep-stationary phase growth, which is likely due to the high expression of \textit{invF} relative to \textit{tnpA} (Figure 5.5D,E,F).

Together, these data show that \textit{tnpA} overexpression affects \textit{invF} levels only when native \textit{tnpA} is expressed at lower levels than \textit{invF}. This suggests the stoichiometry between both transcripts is important, and is consistent with a direct interaction between \textit{tnpA} and \textit{invF}. Additionally, the growth phases where \textit{tnpA} overexpression repressed \textit{sicAsipBC} were the same as those where \textit{tnpA} repressed \textit{invF}, providing additional support to a model where \textit{tnpA} acts through \textit{invF} to repress \textit{sicAsipBC}.

The above results suggested that native expression of \textit{tnpA} does in fact regulate \textit{invF}. We confirmed this by comparing \textit{invF} expression in an IS200 knock-out (\textit{\Delta tnpA\_2/4/6/7}) or WT strain grown to early- or late-exponential phase. In both growth phases, \textit{tnpA} expression was reduced ~2.5-fold in the IS200 knock-out strain, and this correlated with a 2-fold increase in \textit{invF} expression in early-exponential phase and a 1.5-fold increase in \textit{invF} expression in late-exponential phase (Figure 5.6). The smaller effect of reduced \textit{tnpA} expression on \textit{invF} in late-exponential phase is consistent with the above results where \textit{invF} is present at an excess to \textit{tnpA} in this growth phase. This experiment provides strong evidence that IS200 elements play an important role in controlling \textit{invF} expression in \textit{S. Typhimurium}. 

Figure 5.6. Reducing IS200 copy-number increases *invF* expression.

RT-qPCR was performed on WT SL1344 and a derivative in which 4 of 7 copies of IS200 were deleted (*ΔtnpA_2/4/6/7*). RNA was isolated from cells grown to early- or late-exponential phase. Expression of each gene was normalized to the WT strain grown to early-exponential phase. Error bar show the standard error on the mean (n = 4).
5.2.4 Direct interaction between tnpA and invF

We first used IntaRNA (Wright et al, 2014) to find predicted base-pairing interactions between the 5’end of tnpA and invF, and identified a single extended region of complementarity between the first 63nt of tnpA and an interval 104-160 nts upstream of the start codon on invF (Figure 5.7A). This predicted interaction fits with the above data showing that the first 50nt of tnpA is sufficient for repressing invF, and supports tnpA-90 or -110 acting as an sRNA. We used a gel shift assay to determine if tnpA and the 5’end of invF can base-pair in vitro. As the reported transcription start site (+1, TSS) for invF is 132nt upstream of the start codon (in the centre of the predicted pairing region) (Kroger et al, 2012), we elected to start the in vitro transcript for invF at this position. We observed a modest shift in $^{32}$P-labeled invF upon incubation with increasing concentrations of unlabeled tnpA (first 173 nts) (Figure 5.7B, lanes 1-4). Importantly, a complex of the same mobility formed when $^{32}$P-labeled tnpA was incubated with unlabeled invF (lanes 5-8). To determine the specificity of tnpA:invF pairing, we assayed the ability of a previously characterized mutant form of tnpA (tnpA<sup>LS</sup>, (Ellis et al, 2015a)) to pair with invF; pairing was mostly lost as a consequence of the LS mutations to tnpA (lanes 9-11).

We next used Pb<sup>2+</sup> footprinting to define the region on invF that base-pairs with tnpA. $^{5'}$<sup>32</sup>P-labeled invF was incubated with a 5- or 10-fold excess of tnpA (WT or LS) before the addition of Pb(II)-acetate. The most substantial region of pairing was a 7nt interval located 17-23 nts after the invF TSS (lanes 3-5, Figure 5.7C; red letters, Figure 5.7A). Importantly, pairing was lost when tnpA<sup>LS</sup> was incubated with invF (cf. lane 3 to 7), consistent with the gel-shift assay. Interestingly, this region of invF is predicted to base-pair with nts 12-19 of tnpA, which is the location of the ‘A’ processing site on tnpA (Figure 5.3). Accordingly, processing at nt 19 of tnpA would eliminate the seed sequence for the interaction with invF. There was a second region of protection at nts 34-35 on invF, and similar to the main binding site this protection was absent with tnpA<sup>LS</sup>.

To test if this interaction occurs in vivo, we introduced mutations into the tnpA<sub>_7::kan-pTet</sub> construct (T1 mutations) that prevent base-pairing with nts 17-23 of invF (Supplemental Figure S5). We performed RT-qPCR on RNA extracted from
Figure 5.7. Evidence for a base-pairing interactions between invF and tnpA.

(A) Predicted pairing interaction between the first 63 nt of tnpA and a region of invF 104-160 nts upstream of the start codon. Note that the main transcription start site (TSS, +1) for invF is 132 nts upstream of the start codon. invF nucleotides involved in pairing with tnpA are indicated in red; tnpA LS and T1 mutations are shown in bold. (B) Pairing between tnpA and invF was measured by EMSA. $^{32}$P-labeled invF (-132 to +66 relative to the start codon) or tnpA (-103 to +71 relative to the start codon) was incubated with increasing concentrations of unlabeled tnpA or invF respectively (labeled RNA, 2.4 nM; unlabeled RNA 24, 120, 240 nM) and pairing reactions were analyzed by native PAGE. A mutant form of tnpA (tnpA$_{LS}$) was also included in this experiment. Certain lanes have been removed from one gel for clarity (vertical white line separating lanes 8 and 9). Reactions containing only the labeled RNA (lanes 1, 5, and 9) are indicated with ‘-’. (C) Pb$^{2+}$ footprinting was used to analyze base-pairing between 5'$^{32}$P-labeled invF (70 nM) and unlabeled tnpA$_{WT}$ or tnpA$_{LS}$ (same transcripts as in B). An RNase T1 sequencing reaction (G, lane 1) was used to assign positions of lead sensitivity (numbers relative to the 5’end), and an untreated RNA control (UT, lane 2) is shown. Red bars to the right of the gel image highlight tnpA$_{WT}$-dependent protections on invF. (D) RT-qPCR from RNA isolated from the indicated SL1344 strains grown to late-exponential phase (OD$_{600}$ = 1.4). Error bars show standard error on the mean (n = 4).
SL1344 WT, \textit{tnpA\_7::kan-pTet}, and \textit{tnpA\_7::kan-pTet-T1} strains grown to late-exponential phase. Overexpression of the WT \textit{tnpA} reduced \textit{invF} and \textit{sicA} levels 3.5- and 2-fold respectively, while the T1 mutant form of \textit{tnpA} did not affect either of these transcripts (Figure 5.7D). Due to the complex transcriptional regulation of \textit{invF} and the location of the pairing region (~20 nts downstream of the TSS) we have not introduced compensatory mutations to \textit{invF}. However, this experiment showed that the effect of \textit{tnpA} on SPI-1 expression is sequence specific; combined with our \textit{in vitro} pairing experiments, the above data strongly suggests that the 5’ ends of \textit{tnpA} and \textit{invF} base-pair, the consequence of which is reduced \textit{invF} mRNA levels.

5.3 Discussion

In the current work we asked if IS200 encoded transcripts affect gene expression in \textit{S. Typhimurium}. IS200 is an unusual transposon in that it is often present in high copy-number in many \textit{Salmonella} and \textit{Yersinia} spp. but the transposon itself is almost completely dormant. The low transposition frequency of IS200 can be explained by close to no synthesis of the TnpA protein (Ellis et al, 2015a). However, the IS200 transposase mRNA (\textit{tnpA}) is expressed at a moderate level in \textit{S. Typhimurium}, resulting in a paradox where this transposon has evolved to maintain transcription of the transposase mRNA but essentially no translation of the protein. Here we provide an explanation for this paradox by demonstrating that overexpression of \textit{tnpA} alters the expression of at least 73 genes in \textit{S. Typhimurium}, including many genes involved in pathogenesis. We provide evidence that \textit{tnpA} is processed to produce a small regulatory RNA that inhibits expression of the SPI-1 encoded transcription factor \textit{invF} by a base-pairing mechanism.

5.3.1 Ribonucleolytic processing of \textit{tnpA} mRNA generates an sRNA regulator of \textit{invF} expression

We began the current study by profiling the effect of \textit{tnpA} overexpression on gene expression in \textit{S. Typhimurium}. In this experiment we observed strong repression (>2-fold) of 73 genes, 8 of which (\textit{sipC, sipA, sseA, sseL, sigE, sopB, sicA, sipB}) are involved in pathogenesis. Although \textit{tnpA} overexpression also represses art200 expression, four of
these virulence genes (*sicA*, *sipB*, *sipC*, and *sopB*) were repressed by a *tnpA* mutant that is unable to down-regulate art200. As *tnpA* is almost never translated, we speculated that all or part of *tnpA* may act as a non-coding RNA to regulate gene expression in *S. Typhimurium*.

It is now clear that untranslated regions of mRNAs serve as a rich reservoir of sRNAs. As we had observed an effect from overexpressing the 5’portion of *tnpA*, we asked if IS200 expresses a 5’UTR derived sRNA. The typical 5’UTR derived sRNA (5’sRNA) is transcribed from the same promoter as an mRNA and transcription terminates at an intrinsic terminator upstream of the coding sequence for the mRNA (Hershko-Shalev et al, 2016; Loh et al, 2009; Vogel et al, 2003). Although most 5’sRNAs terminate at an intrinsic terminator, post-transcriptional processing occurs for several previously described 5’sRNAs (Bilusic et al, 2014; Papenfort et al, 2015; Vogel et al, 2003). Indeed, our primer extension and Northern analysis revealed that the 5’end of *tnpA* contains two processing sites which produce a ~110nt RNA initiating at the *tnpA* promoter and ending at nt 108 (*tnpA*-110), and a ~90nt species (*tnpA*-90) that is likely generated by processing at nt 19 of *tnpA*-110. Similar to Type II 3’UTR derived sRNAs, the *tnpA* sRNA is likely a stable processing intermediate of *tnpA*, whereby the biogenesis of the *tnpA* sRNA comes as a consequence of ribonucleolytic degradation of an mRNA (Miyakoshi et al, 2015b). Evidence for the instability of *tnpA* 3’ of processing sites comes from the relatively small amount of these downstream products detected by primer extension. At this point we have not identified the ribonuclease responsible for ribonucleolytic processing of *tnpA* but we predict that RNase III and/or RNase E would be involved based on sequence and structural elements at both processing sites. Future work will investigate the precise mechanism of endoribonucleolytic processing of *tnpA* including the potential involvement of art200 in generating *tnpA*-110 and -90.

We found that only the first 50nt of *tnpA* is required for repressing *invF* and *sicAsipBC* expression which fits fully with either *tnpA*-110 or -90 acting as a trans-acting sRNA. Mutations to nt 12-19 on *tnpA* (*tnpAT1*) prevent pairing with *invF in vitro*, and repression of *invF in vivo*. Since *tnpA*-90 lacks the first 19nt, our data points to *tnpA*-110 as the active *tnpA*-derived sRNA, and *tnpA*-90 as an inactivated form of *tnpA*-110. The
second processing event at nt 19 on \textit{tnpA} would therefore represent a point of regulation in \textit{tnpA} mediated repression of \textit{invF}, and ongoing work will address this.

While the current work presents the first example of a bacterial transposon producing a trans-acting sRNA, there are two recent examples of transposase derived sRNAs in archaea. The \textit{Sulfolobus solfataricus} sRNA RNA-257 shares substantial homology with the 3’UTR of the ISC1904 transposase, ORF1182. RNA-257 is believed to be a remnant of transposition reactions, and this sRNA base-pairs with ORF1183, which encodes a putative phosphate transporter. Similar to \textit{tnpA-invF}, base-pairing between RNA-257 and ORF1183 results in degradation of the mRNA (Martens et al, 2013). In \textit{Halobacterium salinarum}, the IS1341 transposase, \textit{tnpB}, expresses more than 10 different sRNAs, one of which regulates growth rate by an undetermined mechanism (Gomes-Filho et al, 2015).

It is perhaps surprising that neither \textit{tnpA}-110 or -90 have been detected in previous work identifying sRNAs in \textit{S. Typhimurium} (Chao et al, 2012; Kroger et al, 2012; Sittka et al, 2008). However, a standard practice in mapping RNA-Seq reads to the reference genome is to omit non-unique reads, and the presence of 7 identical copies of IS200 in SL1344 would result in reads derived from \textit{tnpA} being overlooked. Ambiguity in mapping short reads to multi-copy genes (e.g. many ribosomal proteins) is one limitation to many pipelines for analyzing next-generation sequencing data. Additionally, we found that the most abundant molecule of \textit{tnpA} in WT SL1344 is \textit{tnpA}-90. Since this molecule is processed and possesses a 5’monophosphate, it would be eliminated in protocols designed for dRNA-Seq (e.g. (Kroger et al, 2012)). However, we note that \textit{tnpA} is enriched up to 4.1-fold in Hfq-CoIP experiments (Chao et al, 2012), and the previously characterized Hfq-binding site on \textit{tnpA} (nts 68-83; (Ellis et al, 2015a)) would be present in both \textit{tnpA}-110 and -90. We have not yet investigated the role of Hfq in \textit{tnpA-invF} pairing, in part due to the complications of dysregulated SPI-1 expression (destabilized \textit{hilD}) (López-Garrido et al, 2014) in an \textit{hfq}-null strain.
5.3.2 Regulatory cross-talk between horizontally acquired genes and the S. Typhimurium core genome

*Salmonella* Typhimurium contains a mosaic genome consisting of a core genome complemented with a number of horizontally acquired genetic elements. The core genome is highly conserved amongst Enterobacteriaceae and contains all of the genes required for normal cellular processes. The accessory or ‘flexible’ genome is made up of a number of horizontally acquired genes including pathogenicity islands, prophage, plasmids, and transposons. This flexible genome has been acquired over evolutionary time and provides most of the genes required for virulence (Desai et al, 2013).

Horizontally acquired genes become integrated into host regulatory networks whereby components of the core genome regulate horizontally acquired genes and the core genome itself can be regulated by members of the accessory genome (Lercher & Pál, 2008). In *S. Typhimurium*, the core genome encoded sRNA SgrS represses expression of the SPI-1 effector *sopD* (Papenfort et al, 2012), while the SPI-1 encoded sRNA InvR represses expression of the core genome encoded *ompD* (Pfeiffer et al, 2007). In enterohaemorrhagic *E. coli*, a bacteriophage encoded sRNA, AgvB, represses the core genome encoded sRNA GcvB, thereby increasing expression of many genes involved in amino acid transport (Tree et al, 2014). In the current work we provide evidence of cross-talk between members of the accessory genome, where a transposon derived sRNA controls expression of part of the SPI-1 T3SS by repressing expression of InvF.

Activation of SPI-1 is controlled by a complex network of transcriptional, post-transcriptional, and post-translational regulation (Altier, 2005; Ellermeier & Slauch, 2007). Environmental signals including low oxygen and high osmolarity first converge to activate transcription of *hilD*. HilD, HilC, and RtsA then participate in a feed-forward loop that converges on activation of *hilA*, which in turn activates transcription of structural components of the needle complex (*prg/org*) as well as *invF*, which is a transcriptional activator of secreted effector proteins (*sic/sip* and others) (Figure 5.8). In the current work we have identified a new component of this complex regulation: a processed form of *tnpA* represses *invF* expression in a growth phase dependent manner. We propose that similar to recently identified sRNA ‘sponges’ (Figueroa-Bossi et al,
Expression of virulence genes has an extreme fitness cost for many bacteria including *S. Typhimurium*, *Y. pestis*, and *Shigella flexneri* (Ali et al, 2014; Schuch & Maurelli, 1997; Sturm et al, 2011). For example, single cell analyses revealed that SPI-1 induction dramatically retarded growth and this growth defect was abrogated by deleting the *sic/sip* locus (Sturm et al, 2011). As the *sicA* promoter has the longest relaxation time for SPI-1 encoded genes (Temme et al, 2008), induction of *sic/sip* by InvF represents a key commitment step to virulence and the associated burden of producing effector proteins. We propose a model where repression of *invF* by *tnpA* sets a threshold for *invF* induction that must be passed to induce expression of virulence factors (Figure 5.8). Evidence for this comes from two key experiments. First, when we profiled SPI-1 expression over growth we noted that (i) *tnpA* was expressed higher than *invF* in lag and early-exponential phase, and (ii) overexpression of *tnpA* did not affect *invF* expression in these two growth phases (Figure 5.5). Second, reducing *tnpA* expression two-fold in early-exponential phase increased *invF* expression >2-fold, while the same reduction in *tnpA* expression in late-exponential phase resulted in only a 1.4-fold increase in *invF* (Figure 5.6). This threshold for activation would ensure that InvF is only made once there is a sufficiently high transcriptional activation of the *invF* promoter by HilA. A similar threshold for activation occurs for activation of both *hilD* and *hilA* (Saini et al, 2010; Temme et al, 2008). In the case of HilD, post-translational repression by HilE dampens *hilD* activation, and H-NS repression of *hilA* counteracts transcriptional activation by HilD, HilC, and RtsA (Olekhnovich & Kadner, 2006). An additional role of *tnpA* may be to prevent leaky expression of *invF*, particularly HilC-dependent activation of the alternative promoter for *invF* (p$_{invF}$-2, Figure 8) (Akbar et al, 2003; Lim et al, 2012). While *hilD*, *rtsA*, and *hilA* are strongly repressed prior to induction, *hilC* is expressed at a basal level and expression has minor fluctuations independent of *hilD* (Sturm et al, 2011). As the predicted *tnpA-invF* base-pairing interaction extends 30 nt upstream of the HilA-dependent TSS for *invF*, the HilC-dependent *invF* transcript may be subject to even stronger repression by *tnpA*. 
Figure 5.8. Model for role of tnpA in regulating SPI-1 gene expression.

Environmental signals activate transcription of hilD, which activates a feed-forward loop that converges on activation of hilA. HilA is a transcriptional activator of components of the needle complex (prg/org) as well as invF, a transcriptional activator of SPI-1 effector proteins (sic/sip and other genes). A processed form of tnpA (tnpA-110) base-pairs with invF mRNA and inhibits InvF expression. Processing of tnpA-110 to tnpA-90 would inactivate tnpA and prevent repression of invF. Three activation checkpoints for SPI-1 gene expression are indicated with dashed boxes.
5.3.3 Bacterial transposons as a source of regulatory RNA

Our initial goal for the current work was to determine if any IS200 encoded RNAs affect gene expression in *S. Typhimurium*. Our transcriptomics experiment identified at least 73 genes that are dysregulated by *tnpA* overexpression. We have investigated how *tnpA* impacts on expression of three of these genes (*sicA*, *sipB*, and *sipC*) and our data is consistent with an indirect mechanism where *tnpA* acts through *invF* to control *sicAsipBC* expression. We believe that the effect of *tnpA* on many of the other genes identified here will likewise be indirect, and mediated through a smaller number of direct targets. Regardless of the mechanism by which *tnpA* regulates gene expression in *S. Typhimurium*, we have identified a new way that bacterial transposons can ensure survival: contributing a regulatory RNA. The dual use of a promoter for mRNA and sRNA would ensure that transcription of the transposase is maintained, and post-transcriptional regulation of TnpA expression protects against detrimental effects of transposition.

It is now clear that bacterial sRNAs can be derived from unexpected places. Transposons likely represent an unexplored reservoir of regulatory RNAs that could ultimately provide a benefit to the host organism.

5.4 Materials and Methods

5.4.1 Growth conditions, strains, and plasmids

Unless otherwise stated, *Salmonella* was grown at 37°C with shaking in Lennox Broth (LB; 5 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract). For experiments where RNA was extracted at multiple time-points, overnight cultures were diluted once (1:100 into 7 or 25 mL) and aliquoted (2 mL) into separate culture tubes for each time point. For SPI-1 inducing conditions, cells were grown as previously described (Sittka et al, 2008). For SPI-2 inducing conditions, cells were grown overnight in LB and diluted 1:100 into acidic low-phosphate, low-magnesium (LPM) media (80 mM MES pH 5.8, 5 mM KCl, 7.5 mM (NH₄)₂SO₄, 0.5 mM K₂SO₄, 38 mM glycerol, 0.1% casamino acids [w/v], 8 μM
MgCl₂, 337 µM KH₂PO₄) (Coombes et al, 2004). Where appropriate, antibiotics were used at the following concentrations: tetracycline, 15 µg/mL; chloramphenicol, 20 µg/mL; kanamycin, 25 µg/mL; streptomycin, 150 µg/mL. For experiments with marked alleles, selection was only used in the overnight culture.

All strains and plasmids used in this study are listed in Table S5.1 and oligonucleotides are listed in Table S5.2. S. Typhimurium str. LT2 or SL1344 were considered wild-type strains, and derivative strains were made in the SL1344 background. E. coli DH5α was used for routine cloning and plasmid propagation.

Mutant strains of SL1344 were constructed by Lambda Red recombineering (Datsenko & Wanner, 2000) and all mutations were checked by colony PCR. DBH401 (ΔtnpA_2/4/6/7) was constructed by transducing individual IS200 knockout alleles into a single strain. DBH393 and related strains were created by inserting a kan-pTet cassette in front of tnpA_7 such that the Tet promoter is driving transcription of tnpA (tnpA_7::kan-pTet). Further details of strain and plasmid construction are provided in Supplementary Material.

5.4.2 RNA isolation, Northern blot, and primer extension

Total RNA was prepared by the hot acid phenol method (Aiba et al, 1981). Northern blots were performed as previously described (Ellis et al, 2015b) using 5 or 10 µg of total RNA and 5³²P-labeled oligonucleotide probes (oDH428 tnpA; oDH427, art200) or a uniformly ³²P-labeled riboprobe (5S rRNA, generated with oDH234 and oDH235; art200, generated with oDH450 and oDH394). Primer extension was performed as previously described (Ellis et al, 2015a) using 9 µg of total RNA and primers oDH428 or oDH394. Processed RNA was eliminated by TEX (Epicentre) treatment according to the manufacturer’s instructions.

5.4.3 RNA-Seq and data analysis

S. Typhimurium LT2 was transformed with pDH900 (empty vector), pDH899 (pTet-tnpA_truncWT-255), or pDH914 (pTet-tnpA_truncM1-255). Two colonies from each transformation were each used to inoculate 1 mL of LB-Luria (0.5 g/L NaCl, 10 g/L
tryptone, 5 g/L yeast extract) with tetracycline and were grown for 8 hr. Precultures were subcultured 1:100 into LB-Luria and grown for 16 hr. Total RNA was isolated and treated with TURBO DNase (Ambion) to remove residual genomic DNA and submitted to the London Regional Genomic Centre for library preparation and sequencing. Libraries were prepared with the Ribo-Zero (Gram-Negative Bacteria) (Epicentre) and ScriptSeq v2 (Epicentre) kits. The six libraries (2 biological replicates from each plasmid) were pooled and sequenced with 50 cycles on an Illumina MiSeq. Reads were aligned to the S. Typhimurium LT2 genome (NC_003197) with Rockhopper (McClure et al, 2013) and differential expression was analyzed using ALDEx2 (Fernandes et al, 2014). More detail on data analysis is provided in Supplemental Materials and Methods. Excel files containing RNA-Sequencing reads and ALDEx2 analysis are available as a supplement to this thesis through the Western University Electronic Thesis and Dissertation Thesis Depository.

5.4.4 Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

DNase treated RNA (2 µg) was converted to cDNA with the High-Capacity cDNA Reverse Transcription Kid (Applied Biosystems); cDNA was diluted to 30 ng/µL in TE (50 mM Tris-HCl, pH 8.0, 1 mM EDTA) and stored at -20°C. A minimum of 3 biological replicates were analyzed in technical triplicate in each experiment and the 16S rRNA (rrsA) was used as a reference gene for relative quantitation. Reactions (20 µL) contained 10 ng of cDNA, 500 nM of each primer (Table S5.5), and PowerUP SYBR Green Master Mix (Applied Biosystems). Standard settings on the ViiA 7 Real-Time PCR System were used except for the anneal/extension step, which was performed at 60.5°C. Relative expression of each target was calculated by the efficiency corrected method (Pfaffl, 2001). The amplification efficiency was determined for tnpA (2.20), thrS (2.04), rrsA (2.00), invF (2.12), sipB (2.03), sipC (2.01), and sicA (2.00); an efficiency of 2.0 was used for all other primer pairs.
5.4.5 Western Blot

DBH388 (invF::3X-FLAG-kan) transformed with pDH900 (empty vector), pDH960 (pTet-tpA_trunc-50), or pDH962 (pTet-tpA_trunc-200) was grown to OD600 = 0.5 and cells from 1 mL of culture was collected by centrifugation. For the experiment comparing DBH388 to DBH398 (invF::3X-FLAG-kan tnpA_7::kan-pTet), the volume of culture was adjusted so that an equivalent of 0.5 OD were harvested. The cell pellet was resuspended in 200 µL of SDS sample buffer (60 mM Tris-HCl, pH 6.8, 2% SDS [w/v], 0.01% bromophenol blue [w/v], 1% β-mercaptoethanol [v/v]) and boiled for 5 mins. Samples (10 µL) were resolved on 10% polyacrylamide gels and electroblotted to a PVDF membrane. Membranes were incubated in 5% milk overnight with primary antibody (1:5000 dilution: mouse α-FLAG M2, Sigma; rabbit α-GroES, Sigma; mouse α-DnaK, Enzo), followed by incubation with a 1:5000 dilution of secondary antibody (α-mouse-HRP or α-rabbit-HRP, Promega). Blots were developed with a Pierce ECL 2 Western blotting substrate and a STORM scanner. Membranes were stripped and re-probed for loading controls (GroES/DnaK). Bands were quantitated in ImageQuant and the amount of InvF-3xFLAG was normalized to the internal standard (GroES/DnaK) and then the control strain (empty vector or DBH388).

5.4.6 Electrophoretic Mobility Shift Assay (EMSA) and Lead Footprinting

In vitro pairing experiments were performed as previously described (Ellis et al, 2015a; Ross et al, 2013) except that the RNAs were mixed prior to denaturation.
5.5 Supplemental Material

5.5.1 Supplemental Materials and Methods

5.5.1.1 Strain and plasmid construction

DBH370 (SL1344 ΔhilA::FRT) was created using pKD4 as a template with primers oDH725 and oDH726, and the kanamycin resistance gene was removed by transforming the recombineered strain with pCP20 (a temperature sensitive plasmid expressing FLP recombinase). DBH393 and DBH401 (tnpA_7::kan-pTet and tnpA_7::kan-pTet-T1) were made using pKD4 as a template with one primer containing the pTet promoter and homology to the left end of IS200 (oDH755 or oDH760) and a primer with homology to sequence upstream of tnpA_7 (SL1344_RS179555) (oDH754). The resulting strains have the pTet promoter driving constitutive transcription of the full-length tnpA mRNA initiating at nucleotide 46 of IS200 and an upstream kanamycin resistance gene expressed in the opposing direction (see Figure S5.3B). DBH398 was made by first constructing a strain identical to DBH393 except using pKD3 as a template. The invF::3X-FLAG-kan allele from DBH388 was then introduced to this strain by P22 transduction.

The IS200 knock-out strain was constructed by sequential P22 transduction of one strain with individual IS200 deletions. First, a pool of IS200 deletions was made by constructing a universal IS200 knock-out cassette (able to target any of the 7 identical copies of IS200 in SL1344) using either pKD4 (kanR cassette) or pKD3 (cmR cassette) with primers oDH613 and oDH614. Individual colonies from both the kanR and cmR pool were screened by PCR to determine which copy of IS200 was disrupted. Next, P22 (HT-105/int-201) transduction was used to move two marked alleles (ΔtnpA_2::cm, SL1344_RS06060; ΔtnpA_7::kan, SL1344_RS179555) sequentially to DBH347 (WT SL1344). After purification of transductants on EBU plates (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCL, 0.25% glucose [w/v], 5 mM K2HPO4, 0.00125% Evans Blue [w/v], 0.0025% uranine [w/v], 15 g/L agar), colonies were cross-streaked against P22 H5 to select phage-sensitive colonies. Both selectable markers were removed by transforming the strain with pCP20 and both alleles were checked by colony PCR. P22 transduction was then repeated to move ΔtnpA_6::cm (SL1344_RS17020) and ΔtnpA_4::kan.
(SL1344_RS09790) to create DBH400 (SL1344 ΔtnpA_2::FRT ΔtnpA_7::FRT ΔtnpA_6::cm ΔtnpA_4::kan).

Plasmids overexpressing truncated forms of tnpA were modified from the pACYC184-derived pDH899 (pTet-tnpA_{trunc}^{WT}-255) (Ellis et al, 2015a). A PCR product amplified from DBH199 (WT LT2) genomic DNA with oDH531 and oDH746 (pTet-tnpA_{trunc}-50), oDH743 (pTet-tnpA_{trunc}-200), or oDH742 (pTet-tnpA_{trunk}-250) was digested with EcoRI and cloned into the same site of pACYC184.

5.5.1.2 RNA-Seq Data Analysis

Transcripts predicted by Rockhopper were manually annotated with the genomic interval and whether it was intergenic (predicted RNA, e.g. 4824999-4825049_{predicted RNA}) or intragenic (antisense, e.g. 1906043-1906092_{antisense: fadR}) prior to expression analysis.

Differential expression was analyzed using ALDEx2 version 1.0.0 (Fernandes et al, 2014) which is a tool for analyzing the compositional nature of high-throughput sequencing data. Since ALDEx2 compares two conditions this analysis was performed on each pairwise comparison (i.e. Vector vs pTet-tnpA_{trunc}^{WT}-255, Vector vs pTet-tnpA_{trunc}^{M1}-255, pTet-tnpA_{trunc}^{WT}-255 vs pTet-tnpA_{trunk}^{M1}-255). Briefly, the relative expression (abundance) of each gene within a sample was calculated as the median centered log-ratio (clr) from 1000 Monte-Carlo Dirichlet instances. The ALDEx2 effect size is then calculated for each feature as the median log2 difference between conditions (i.e. strains) divided by the maximum difference within a condition (i.e. difference in relative abundance between biological replicates). Accordingly, the effect size is a measure of the confidence one has in the difference in expression for a gene between two conditions, but does not necessarily inform on the magnitude of the difference. An effect size of 2 means that the difference between conditions (i.e. log2 fold-change between two strains) is twice as large as the greater difference within a condition (variance between biological replicates). We took a more conservative approach for considering genes differentially expressed and used an effect size of 2 as the cutoff for differential expression.
5.5.2 Supplemental Tables and Figures
### Table S 5.1. Strains and plasmids for Chapter 5.

<table>
<thead>
<tr>
<th>Strains and plasmids</th>
<th>Relevant Genotype/ Description</th>
<th>Reference/source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S. Typhimurium</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBH199</td>
<td>Wild-type LT2, Str&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Miguel Valvano</td>
</tr>
<tr>
<td>DBH347</td>
<td>Wild-type SL1344, Str&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Salmonella Genetic Stock Centre (SGSC438)</td>
</tr>
<tr>
<td>DBH370</td>
<td>DBH347 ΔhilA::FRT, Str&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This Study</td>
</tr>
<tr>
<td>DBH388</td>
<td>JVS-2325 (SL1344 invF::3X-FLAG-kan), Str&lt;sup&gt;R&lt;/sup&gt;Kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Jörg Vogel</td>
</tr>
<tr>
<td>DBH393</td>
<td>DBH347 tnpA&lt;sub&gt;7&lt;/sub&gt;::pTet-kan, Str&lt;sup&gt;R&lt;/sup&gt;Kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This Study</td>
</tr>
<tr>
<td>DBH398</td>
<td>DBH347 tnpA&lt;sub&gt;7&lt;/sub&gt;::pTet-cm invF::3X-FLAG-kan, Str&lt;sup&gt;R&lt;/sup&gt;Kan&lt;sup&gt;R&lt;/sup&gt;Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This Study</td>
</tr>
<tr>
<td>DBH400</td>
<td>DBH347 ΔtnpA&lt;sub&gt;_2&lt;/sub&gt;::FRT ΔtnpA&lt;sub&gt;_7&lt;/sub&gt;::FRT ΔtnpA&lt;sub&gt;_6&lt;/sub&gt;::cm ΔtnpA&lt;sub&gt;_4&lt;/sub&gt;::kan</td>
<td>This Study</td>
</tr>
<tr>
<td>DBH401</td>
<td>DBH347 tnpA&lt;sub&gt;_7&lt;/sub&gt;::pTet-kan-T1 Str&lt;sup&gt;R&lt;/sup&gt;Kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This Study</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pDH739</td>
<td>pCP20, temperature sensitive FLP expression plasmid, Ap&lt;sup&gt;R&lt;/sup&gt;Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Datsenko &amp; Wanner, 2000)</td>
</tr>
<tr>
<td>pDH742</td>
<td>pKD4, kanamycin cassette for Lambda Red recombineering, Ap&lt;sup&gt;R&lt;/sup&gt;Kn&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Datsenko &amp; Wanner, 2000)</td>
</tr>
<tr>
<td>pDH743</td>
<td>pKD3, chloramphenicol cassette for Lambda Red recombineering, Ap&lt;sup&gt;R&lt;/sup&gt;Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Datsenko &amp; Wanner, 2000)</td>
</tr>
<tr>
<td>pDH899</td>
<td>pACYC184-derived plasmid expressing tnpA&lt;sub&gt;trunc&lt;/sub&gt;&lt;sup&gt;WT&lt;/sup&gt;-255 from the pTet promoter, Cm&lt;sup&gt;S&lt;/sup&gt;Tet&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Ellis et al, 2015a)</td>
</tr>
<tr>
<td>pDH900</td>
<td>pACYC184-derived empty vector control, Cm&lt;sup&gt;S&lt;/sup&gt;Tet&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Ellis et al, 2015a)</td>
</tr>
<tr>
<td>pDH914</td>
<td>pACYC184-derived plasmid expressing tnpA&lt;sub&gt;trunc&lt;/sub&gt;&lt;sup&gt;M1&lt;/sup&gt;-255 from the pTet promoter, Cm&lt;sup&gt;S&lt;/sup&gt;Tet&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Ellis et al, 2015a)</td>
</tr>
<tr>
<td>pDH960</td>
<td>pACYC184-derived plasmid expressing tnpA&lt;sub&gt;trunc&lt;/sub&gt;50 from the pTet promoter, Cm&lt;sup&gt;S&lt;/sup&gt;Tet&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This Study</td>
</tr>
<tr>
<td>pDH962</td>
<td>pACYC184-derived plasmid expressing tnpA&lt;sub&gt;trunc&lt;/sub&gt;200 from the pTet promoter, Cm&lt;sup&gt;S&lt;/sup&gt;Tet&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This Study</td>
</tr>
<tr>
<td>pDH963</td>
<td>pACYC184-derived plasmid expressing tnpA&lt;sub&gt;trunc&lt;/sub&gt;250 from the pTet promoter, Cm&lt;sup&gt;S&lt;/sup&gt;Tet&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This Study</td>
</tr>
<tr>
<td>Name</td>
<td>Sequence (5’ to 3’)</td>
<td>Description/Use</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>oDH725</td>
<td>TATAACTTTTCAACCCTGTAAGAGAATACACTAT TATCATGGTGATAGGGCTGGAGCTGCTCC</td>
<td>Amplifying kan-cassette from pKD4, <em>hilA</em> deletion (DBH370)</td>
</tr>
<tr>
<td>oDH726</td>
<td>TCTCCCTCTCTCAGATTAAAAAGCCGCTTTAA AGCGGGCATATGAAATATCTCCTCCCTAG</td>
<td></td>
</tr>
<tr>
<td>oDH727</td>
<td>GCACCAGATATACGCGATGTCCT</td>
<td>Checking <em>hilA</em> locus</td>
</tr>
<tr>
<td>oDH728</td>
<td>GCCGCGAGTTTTATCAGCGG</td>
<td></td>
</tr>
<tr>
<td>oDH754</td>
<td>TACGAGGATATATATATATATGTTTTATCAAAGCCGCTTTAA AGCGGGCATATGAAATATCTCCTCCCTAG</td>
<td>Amplifying kan-cassette from pKD4, <em>tnpA_7::kan-pTet</em> strain (DBH393, DBH398)</td>
</tr>
<tr>
<td>oDH755</td>
<td>TAACAAATACAAAGGGTTTTATAAAGCTCGTTTAAAGCGCTTTAAAGCTCGTTA AGCGGGCATATGAAATATCTCCTCCCTAG</td>
<td></td>
</tr>
<tr>
<td>oDH760</td>
<td>CAGACCCGAAGATTTAAAAATCTAAAAAGCCGCTTTAAAGCGGGCATATGAAATATCTCCTCCCTAG</td>
<td>Amplifying kan-cassette from pKD4, <em>tnpA_7::kan-pTet-T1</em> (DBH401)</td>
</tr>
<tr>
<td>oDH613</td>
<td>TGCTATGAGAAACCCCAGCCTAGGGCTGGGTT TCCGAAGTTGATAGGGCTGGGCTGCTTC</td>
<td>Amplifying kan-cassette from pKD3/4, <em>ΔJS200::cm/kan</em> (DBH400)</td>
</tr>
<tr>
<td>oDH614</td>
<td>CATGTGAATATGCTGACAACATTACTGCTGCTTC CAGCATATGAAATACCTCCTCTTA</td>
<td></td>
</tr>
<tr>
<td>oDH615</td>
<td>GACGTATGACTTCCAGAGTGTC</td>
<td>Checking <em>tnpA_1</em> locus (SL1344_RS04595)</td>
</tr>
<tr>
<td>oDH616</td>
<td>GCCCTTATCCGCTTACCGG</td>
<td>Checking <em>tnpA_2</em> locus (SL1344_RS06060)</td>
</tr>
<tr>
<td>oDH617</td>
<td>GAGCTGACGGCTAAGGCG</td>
<td>Checking <em>tnpA_3</em> locus (SL1344_RS09075)</td>
</tr>
<tr>
<td>oDH618</td>
<td>CCGCATATCATAAAACGTCG</td>
<td>Checking <em>tnpA_4</em> locus (SL1344_RS09790)</td>
</tr>
<tr>
<td>oDH619</td>
<td>CGACGAGAAGCTGCTGACG</td>
<td>Checking <em>tnpA_5</em> locus (SL1344_RS12685)</td>
</tr>
<tr>
<td>oDH620</td>
<td>CATAGTGTTTCTATAGGGCTGAGG</td>
<td>Checking <em>tnpA_6</em> locus (SL1344_RS17020)</td>
</tr>
<tr>
<td>oDH621</td>
<td>CAGCCCAAGATTTAAAAATCTAAAAAGCCGCTTTAAAGCGGGCATATGAAATATCTCCTCCCTAG</td>
<td>Checking <em>tnpA_7</em> locus (SL1344_RS17955)</td>
</tr>
<tr>
<td>oDH622</td>
<td>CTGCTGCTGCTTCCATATGGAGG</td>
<td></td>
</tr>
<tr>
<td>oDH623</td>
<td>TTGTCATGAGCTGACGAGG</td>
<td></td>
</tr>
<tr>
<td>oDH624</td>
<td>GAGCTGACGGCTAAGGCG</td>
<td></td>
</tr>
<tr>
<td>oDH625</td>
<td>CCGCATATCATAAAACGTCG</td>
<td></td>
</tr>
<tr>
<td>oDH626</td>
<td>CGTACATCTGGTCTGCTACC</td>
<td></td>
</tr>
<tr>
<td>oDH627</td>
<td>AAGAATTCTCATGTGAGCTGAGCTTGATTTCACTAGG</td>
<td>Forward primer for pTet-tnpA&lt;sub&gt;trunc&lt;/sub&gt;</td>
</tr>
<tr>
<td>oDH628</td>
<td>TCAGCTGAGCTGAGCTGAGC</td>
<td>Reversible primer for pTet-tnpA&lt;sub&gt;trunc&lt;/sub&gt;-50 (pDH690)</td>
</tr>
<tr>
<td>oDH629</td>
<td>AGCAGAATCTGAGCTGAGG</td>
<td>Reversible primer for pTet-tnpA&lt;sub&gt;trunc&lt;/sub&gt;-200 (pDH692)</td>
</tr>
<tr>
<td>oDH742</td>
<td>TGGACGAAATTCAAAAAACCAGCAGGTATAATCTGCTGGCGTACGTTTTTCCATTCAC</td>
<td>Reverse primer for pTet-tnpA&lt;sub&gt;trunc&lt;/sub&gt;-250 (pDH963)</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------------------------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td><strong>Northern blot and <em>in vitro</em> transcription</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>oDH428</td>
<td>GTTGCCAGACCAGGAAGATG</td>
<td>Probe for <em>tnpA</em></td>
</tr>
<tr>
<td>oDH427</td>
<td>CATCTTGCGGCTCTGGCAAC</td>
<td>Probe for art200</td>
</tr>
<tr>
<td>oDH234</td>
<td>TAATACGACTCACTATAGGCCTGGCATCCCTACTTCG</td>
<td>ITR template for 5S rRNA Northern probe</td>
</tr>
<tr>
<td>oDH235</td>
<td>CGGCGTAGCGGCCTGTG</td>
<td>Forward primer for <em>tnpA</em>&lt;sup&gt;WT/LS&lt;/sup&gt; ITR template, contains T7 promoter sequence</td>
</tr>
<tr>
<td>oDH450</td>
<td>TAATACGACTCACTATAGGCTTCAGCTTTAAGCCAG</td>
<td>Reverse primer for <em>tnpA</em> ITR template</td>
</tr>
<tr>
<td>oDH394</td>
<td>TATTTGGGCGCCAGAAAAACTATG</td>
<td></td>
</tr>
<tr>
<td>oDH749</td>
<td>TAATACGACTCACTATAGGGATAAAAATGATTCTTACTGGGTGC</td>
<td>Forward primer for <em>invF</em>&lt;sup&gt;WT&lt;/sup&gt; ITR template, contains T7 promoter sequence</td>
</tr>
<tr>
<td>oDH750</td>
<td>GCAAAAAAGCAGCGGCCTTCTCCTG</td>
<td>Reverse primer for <em>invF</em> ITR template</td>
</tr>
<tr>
<td><strong>RT-qPCR</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>oDH704</td>
<td>TTCAGTTGCGATACATGGCCGC</td>
<td><em>sicA</em></td>
</tr>
<tr>
<td>oDH705</td>
<td>GAAAGCCGCCACGCTTAAAGA</td>
<td><em>sipB</em></td>
</tr>
<tr>
<td>oDH706</td>
<td>TGAGACAGCGAAACATCGCC</td>
<td><em>sipC</em></td>
</tr>
<tr>
<td>oDH707</td>
<td>GGAACAAAGTCCCGCCGAGAG</td>
<td><em>rrsA</em> (16S rRNA)</td>
</tr>
<tr>
<td>oDH702</td>
<td>TGCCGTGCTGTTCGATGCTCAT</td>
<td><em>thrS</em></td>
</tr>
<tr>
<td>oDH703</td>
<td>AAACCGAGTTACGCGCAGAG</td>
<td><em>invF</em></td>
</tr>
<tr>
<td>oDH661</td>
<td>TCTGGGAAACTGCTGTATGGAGG</td>
<td><em>tnpA</em></td>
</tr>
<tr>
<td>oDH662</td>
<td>CGTAGGGATCTGGAGCTTGCTTC</td>
<td><em>prgH</em></td>
</tr>
<tr>
<td>oDH686</td>
<td>TCACTCTTTGTGCGCATCTGT</td>
<td></td>
</tr>
<tr>
<td>oDH687</td>
<td>TGCCTCAAGATCAACGTCA</td>
<td></td>
</tr>
<tr>
<td>oDH710</td>
<td>GTCTCCCTGATACGTGTCG</td>
<td></td>
</tr>
<tr>
<td>oDH711</td>
<td>AGCCCGGAGCATGTTTAT</td>
<td></td>
</tr>
<tr>
<td>oDH394</td>
<td>TATTTGGGCGCCAGAAAAACTATG</td>
<td></td>
</tr>
<tr>
<td>oDH428</td>
<td>GTTGGCAGACCAGGAAGATG</td>
<td></td>
</tr>
<tr>
<td>oDH764</td>
<td>GCCAGCTGGGATATAAGGT</td>
<td></td>
</tr>
<tr>
<td>oDH765</td>
<td>TGCGCTGGCGCAAAAATGAAA</td>
<td></td>
</tr>
</tbody>
</table>
Figure S 5.1. Growth phase dependent regulation of art200 and art200 knockdown by tnpA\textsubscript{trunc} overexpression.

(A) An overnight culture of wild-type SL1344 (DBH347) was subcultured 1:100 into Lennox broth and grown for 7.5hrs with shaking. The absorbance at 600 nm (OD\textsubscript{600}) of 2 mL aliquots of culture was measured regularly over this period. RNA was extracted from cells grown to early-exponential (OD\textsubscript{600}=0.2), mid-exponential (OD\textsubscript{600}=0.5), late-exponential (OD\textsubscript{600}=1.2), early-stationary (OD\textsubscript{600}=2.0) and 2 or 4h after cells reached stationary phase (2.0+2h and 2.0+4h). Cells were also subcultured into SPI-2 inducing media (LPM salts, pH 5.8) or grown overnight under SPI-1 inducing conditions (overnight growth in a tightly sealed falcon tube containing LB with a final NaCl concentration of 0.3M) to an OD\textsubscript{600} of 0.35 or 0.9 respectively. (B) Art200 levels were measured by Northern blot from RNA collected from the different growth conditions in (A). The 5S rRNA was used as a loading control. (C) Northern blot showing the effect of tnpA\textsubscript{trunc} expression on art200. S. Typhimurium LT2 was transformed with an empty vector (pDH900) or plasmid overexpressing tnpA\textsubscript{trunc}\textsubscript{WT}-255 (pDH899) or tnpA\textsubscript{trunc}\textsubscript{M1}-255 (pDH914) and RNA was prepared from cultures grown to deep stationary phase. The blot was probed successively for art200, tnpA, and the 5S rRNA as a loading control. Note that these samples were ultimately used for RNA-Seq.
Figure S 5.2. Processing of *tnpA*.

(A) Processing at site “B” generates a 5’monophosphate on *tnpA*105/118 molecules. RNA isolated from WT or *tnpA_7::kan-pTet* (pTet (WT)) cells was treated with 5’monophosphate-dependent terminator exonuclease (TEX, +) or left untreated (-) and analyzed by primer extension using oDH394 (anneals to *tnpA* coding sequence). L, C, and A, refer to a DNA size ladder and ddNTP sequencing lanes respectively. (B) Secondary structure of *tnpA* highlighting processing sites on *tnpA* determined by primer extension (Figure 5.3). Processing at only the ‘purple’ site (nts 105, 107, 108) would yield the ~110nt species detected by Northern blot, while processing at the ‘red’ (nt 19) and ‘purple’ sites would produce the 90 nt species in the Northern blot. The sequence on *tnpA* recognized by the Northern probe is indicated in green. Numbering is relative to the
transcription start site of *tnpA* expressed from the pTet promoter. (C) Proposed processing pathway for *tnpA*. Full-length *tnpA* would be processed at site ‘B’ (purple) generating *tnpA*-110 and *tnpA*-111-? (3’ end unknown). This 3’ fragment is inferred to be unstable and presumably degraded by cellular RNases. Subsequent processing at site “A” on *tnpA* (red) generates *tnpA*-90, which is the most stable *tnpA* species. The binding sites for the primers used to detect *tnpA* in RT-qPCR experiments are shown in green. Accordingly, only unprocessed *tnpA* would be detected in these experiments. (D) Northern blot of RNA isolated from cells expressing various *tnpA* trunc constructs from a plasmid. The processed forms of *tnpA* were detected with an oligonucleotide probe that anneals to the 5’UTR of *tnpA* (indicated in (B)). Note that the probe used for detecting *tnpA* anneals to sequence absent (*tnpA* trunc-50) or mutated (*tnpA* truncM1-255) in certain constructs, and accordingly only endogenous *tnpA* was detected. The membrane was stripped and reprobed first for art200 and then 5S rRNA as a loading control.
Figure S 5.3. Schematic of *tnpA* overexpression constructs used in this work.

(A) Schematic of various *tnpA* trunc constructs used in this work. The WT or M1 *tnpA* trunc-255 constructs (used in RNA-Seq) consist of the first 255 nts of *tnpA* fused to the 3’ end of SgrS (nts 143-227, indicated in red). For subsequent constructs (*tnpA* trunc -50, -200, -250), only the SgrS Rho-independent transcriptional terminator was included. Numbering for all *tnpA* constructs indicates how many nucleotides of *tnpA* (beginning with the transcription start site) are included, and all constructs are expressed from the pTet promoter on a pACYC184-derived vector. The secondary structure of the *tnpA* 5’UTR was determined previously (Ellis et al. 2015a). (B) Schematic of the native *tnpA* _7_ locus (in WT SL1344, DBH347) and *kan*-pTet cassette inserted to drive *tnpA* expression from the pTet promoter (DBH393 and DBH401). The 7th copy of IS200 (*tnpA* _7_, SL1344_RS17955; numbering of IS200 elements is according to chromosomal coordinates) is flanked by *bigA* and *yhfL*. The cassette to overexpress *tnpA* was amplified from pKD4 and includes a kanamycin resistance gene (blue) flanked by FLP recombinase target sites (FRT) and the pTet promoter driving expression of *tnpA* beginning at nt 45 of IS200. The end of this cassette is ~300bp away from *yhfL*. 
Figure S 5.4. Growth phase dependent regulation of \textit{invF} by \textit{tnpA}.

Lennox broth (2 mL) was inoculated with single colonies (4 biological replicates) of WT or \textit{tnpA\_7::kan-pTet} cells and RNA was harvested after 18 hrs (\textbf{F}, deep-stationary phase, $OD_{600}=1.80$). The 18hr cultures were used to seed subcultures, and RNA was isolated after 1.25 hrs (\textbf{B}, lag-phase, $OD_{600}=0.10$), 2 hrs (\textbf{C}, early-exponential phase, $OD_{600}=0.37$), 3hrs (\textbf{D}, late-exponential phase, $OD_{600}=1.28$), or 4hrs (\textbf{E}, early-stationary phase, $OD_{600}=2.01$). The $OD_{600}$ for each time point and strain is plotted in (\textbf{A}), with error bars showing standard error on the mean. Expression of the indicated genes was measured by RT-qPCR; the $\Delta Ct$ (relative to 16S rRNA, \textit{rrsA}) is reported to compare expression between different genes and growth phases. Error bars show the standard error of the mean of 4 biological replicates, each measured in technical triplicate.
Figure S 5.5. The T1 mutations to \textit{tnpA} disrupt base-pairing with \textit{invF} \textit{in vitro}.

Lead(II) acetate footprinting was used to analyze base-pairing between 5’\textsuperscript{32}P-labeled invF (70 nM) and unlabeled \textit{tnpA}\textsuperscript{WT} or \textit{tnpA}\textsuperscript{T1} (700 nM; see also Figure 5.7). An RNase T1 sequencing reaction (G) was used to assign positions of lead sensitivity (numbers relative to the 5’end of invF), and an untreated RNA control (UT) is shown. Red bars to the right of the gel image highlight \textit{tnpA}\textsuperscript{WT} dependent protections on \textit{invF}.
5.6 References


Siguier P, Gourbeyre E, Chandler M (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS microbiology reviews* **38**: 865-891


van Nues Rob W, Castro-Roa D, Yuzenkova Y, Zenkin N (2016) Ribonucleoprotein particles of bacterial small non-coding RNA IsrA (IS61 or McaS) and its interaction with RNA polymerase core may link transcription to mRNA fate. Nucleic acids research 44: 2577-2592


Chapter 6

6 General Discussion

Mobile genetic elements such as transposons exist in a delicate relationship with the host organism. Transposons must mobilize at a frequency high enough to ensure survival of the element but below a level that would damage or kill the host cell. Bacteria tolerate low levels of transposition and benefit from rare transposon-induced genetic changes that can provide adaptive advantages and aid genome evolution, and can also directly benefit from transposon-encoded proteins. This balance is reached through multiple levels of regulation achieved by the cooperation of transposons and their bacterial hosts. Most transposons employ a molecular form of self-restraint where the transposon itself has evolved to transpose at a low frequency. This is achieved largely by poor expression of the transposase protein and includes low levels of transcription and inefficient translation initiation. Additionally, transposition reactions are inherently inefficient and transposons encode regulatory RNAs and/or proteins that act to inhibit transposition. The host organism further limits transposon activity by regulating transcription initiation as well as steps in transposition that occur after the transposase protein is synthesized. In general, the frequency of transposition is dependent on the amount of transposase protein synthesized and accordingly transposase expression is a major point of regulation (Gueguen et al, 2005; Nagy & Chandler, 2004; Siguier et al, 2014; Siguier et al, 2015).

Given the importance of regulatory RNA in bacteria (Bouloc & Repoila, 2016; Papenfort & Vogel, 2014; Storz et al, 2011; Wagner & Romby, 2015), post-transcriptional regulation of bacterial transposase proteins could be an unappreciated wide-spread mechanism for host-control of transposons. In this thesis I’ve shown that the RNA-binding protein Hfq—a core component of post-transcriptional regulation in bacteria—regulates IS10 and IS200 transposase expression by two distinct mechanisms. The finding that a transposase mRNA (IS200) can regulate host gene expression provides an unexpected twist on our understanding of the host-transposon relationship in bacteria. My work has provided novel insight into the function of Hfq and suggests that RNA transactions are an important component of host-transposon interactions.
6.1 Host-transposon interactions at the RNA level

6.1.1 Hfq represses IS10 transposase expression by two distinct mechanisms

The serendipitous discovery that Hfq represses IS10/Tn10 transposition provided the first example of a host-protein controlling transposase expression at the post-transcriptional level (Ross et al, 2010). This work concluded that Hfq repressed RNA-IN translation by antisense-dependent and -independent mechanisms. For the former, Hfq accelerated pairing between RNA-OUT and RNA-IN 8-fold, suggesting that Hfq could catalyze antisense regulation as described for sRNAs.

Work is presented in Chapter 2 that follows up on this initial finding and shows that the distal surface of Hfq binds RNA-IN while the proximal surface interacts with RNA-OUT. Overall, Hfq treats the asRNA and mRNA components of the IS10 antisense system like a canonical sRNA/mRNA pair. A U-rich internal bulge on RNA-OUT engages the proximal surface of Hfq, in contrast to most sRNAs where the proximal surface binds the poly-U tail and stem-loop structure of Rho-independent terminators. It therefore seems that Hfq employs a single RNA-binding surface to recognize different motifs on functionally related RNAs. Antisense RNAs were presumed to be Hfq-independent because of the extensive complementarity with their target mRNA and it therefore was not clear why Hfq would be needed in the IS10 system (Brantl, 2007; Thomason & Storz, 2010). IS10 antisense pairing involves 35 bp of perfect complementarity, however previous genetic evidence and secondary structure predictions suggested that RNA-OUT formed a single stem-loop structure where 24 nts involved in antisense pairing were engaged in intramolecular base-pairing (Case et al, 1989; Simons & Kleckner, 1983). Initial models for RNA-IN/RNA-OUT pairing relied on the assumption that the 5’ end of RNA-IN was unstructured. The unstructured 5’ end of RNA-IN interacts with the terminal loop of RNA-OUT and full pairing was believed proceed through a strand-displacement reaction where intramolecular base-pairs on RNA-OUT were exchanged for intermolecular base-pairs with RNA-IN (Brantl, 2002; Kittle et al, 1989; Kleckner, 1990; Wagner et al, 2002). My structure probing experiments show that the 5’ end of RNA-IN is moderately structured (18 nts involved in pairing are involved in
intramolecular base-pairs), and this would mean that full RNA-IN/RNA-OUT pairing would actually involve a double strand-displacement reaction (i.e. intramolecular base-pairs on both RNA-IN and RNA-OUT are exchanged for intermolecular base-pairs between RNA-IN and RNA-OUT). I propose that RNA restructuring by Hfq would bypass the strand-displacement step and this accelerates the rate of full pairing (Figure 6.1A). This model predicts that more extensive IN-OUT pairing occurs in the presence of Hfq and footprinting experiments with $^{32}$P-labeled RNA-OUT show that this is the case (data not shown).

Exactly how Hfq alters secondary structure is not clear although new insight into the elusive role of the Hfq CTD may provide some clues. RNA-IN interacts with both the distal and proximal surfaces of Hfq which is in contrast to many mRNAs that interact solely with the distal surface (Sobrero & Valverde, 2012; Updegrove et al, 2016). However, an interaction between the proximal surface of Hfq and an mRNA appears to be required for sRNA-mRNA pairs where the 5’end of an mRNA is structured (e.g. rpoS, (Henderson et al, 2013; Soper et al, 2011); fhlA (Salim & Feig, 2010); glmS, (Salim et al, 2012); sodB, (Geissmann & Touati, 2004)) (Beisel et al, 2012; Schu et al, 2015; Zhang et al, 2013). Recent work suggests that the Hfq CTD antagonizes sRNA binding to the Hfq by ‘sweeping’ the proximal surface and promoting dissociation of the RNA from this binding site (Santiago-Frangos et al, 2016). An mRNA like RNA-IN would be stably engaged with Hfq through the distal surface and interactions between the CTD, proximal surface, and the mRNA may promote unfolding. In this way the identification of proximal surface binding sites on a mRNA may be an indicator that restructuring is employed by Hfq to promote pairing with either an sRNA or asRNA.
Figure 6.1. Summary of new insights into post-transcriptional control of transposons IS10 and IS200 gained in this work.

(A) Overview of IS10 antisense pairing in the (i) absence or (ii) presence of Hfq. Hfq (green circles) would not affect the initial interaction between RNA-IN (red) and RNA-OUT, but complete pairing without Hfq would require a presumably slow double-strand displacement reaction to occur for full pairing. Hfq remodels the secondary structure of both RNA-IN and RNA-OUT to expose pairing sequences which accelerates full pairing.

(B) Control of translation initiation for (i) IS10 and (ii) IS200 transposase. (i) 30S ribosome binding to RNA-IN is inhibited by RNA-OUT and Hfq binding to the Shine-Dalgarno sequence (SD). (ii) 30S ribosome binding is inhibited by secondary structure in the *tnpA* 5′UTR that includes the SD. A putative translational enhancer sequence just
upstream of the SD may be recognized by ribosomal protein S1 which could tether the 30S ribosome to \textit{tnpA} or promote RNA modelling to expose the SD and start codon (AUG). Both Hfq and art200 (purple) block this translational enhancer sequence and together with secondary structure ensure that translation initiation on \textit{tnpA} does not occur.
In Chapter 3 I investigated the antisense-independent role of Hfq in repressing IS10 transposase translation and show that Hfq binding to the TIR of RNA-IN is sufficient for inhibiting translation (Figure 6.1B, i). The dispensability of sRNA-binding surfaces, \textit{in vitro} toeprinting experiments, and the correlation between \textit{in vitro} and \textit{in vivo} experiments measuring Hfq-binding to RNA-IN and the effect on translation all support an sRNA-independent regulatory role for Hfq. Notably, this is only the second example of Hfq inhibiting translation fully-independent of an sRNA. In the previous example, Hfq is proposed to autoregulate its expression by binding to the TIR of \textit{hfq} mRNA although in this case the involvement of an sRNA was not conclusively ruled out \textit{in vivo} (Vecerek et al, 2005). An important caveat to this form of regulation is that a sustained interaction between Hfq and the target mRNA is required for translational control. We tested this by inducing other Hfq-binding sRNAs to ask if \textit{in vivo} competition for Hfq binding would lead to derepression of RNA-IN translation. We identified a single sRNA (ChiX) that was able to activate IS10 transposase expression by preventing Hfq-binding to RNA-IN. I proposed that this was due to ChiX interacting with the distal surface of Hfq which is an atypical Hfq-sRNA interaction. Subsequent work has supported this proposal and it seems that ChiX is a member of a small class of sRNAs (termed Class II sRNA) that bind the distal surface of Hfq (Malecka et al, 2015; Santiago-Frangos et al, 2016; Schu et al, 2015; Updegrove et al, 2016).

We have suggested for several years that Hfq could link transposition to stress where induction of sRNA would titrate Hfq away from transposase mRNA and this would increase translation and subsequently transposition. The characterization of the ‘direct’ pathway used by Hfq to inhibit translation provides a clear mechanism for how this could occur and my experiments with ChiX are proof of this principle. The fact that only Class II sRNAs could titrate Hfq away from RNA-IN might ensure that transposition is not readily activated in response to other stress-induced sRNAs, the majority of which belong to Class I (bind proximal surface only). Alternatively, many stresses induce multiple sRNA response pathways and the simultaneous accumulation of many Class I sRNAs might lead to transposon activation similar to what we observed for a single Class II sRNA.
6.1.2 Cooperation between host- and transposon-encoded factors controls transposase translation

My work investigating how Hfq enhances antisense pairing in the IS10 system prompted me to ask if other transposons would be subject to similar regulation. An antisense RNA was previously identified in the IS30 system although in this case pairing occurs within the transposase coding sequence (Arini et al, 1997). One of the first attempts to identify novel sRNAs in Salmonella Typhimurium used an Hfq immunoprecipitation (Hfq-IP) followed by RNA-Seq of bound RNAs (Sittka et al, 2008). This study identified an Hfq-binding sRNA (art200) antisense to the 5’ UTR of the IS200 transposase mRNA (tnpA) and it therefore seemed likely that Hfq would regulate IS200 similarly to IS10. I found that art200 represses tnpA translation similar to RNA-IN/OUT in IS10, but in this case Hfq does not facilitate pairing. However, Hfq does directly inhibit translation initiation on tnpA, albeit to a lesser extent than with RNA-IN. Hfq, art200, and an intrinsic secondary structure formed in the tnpA 5’ UTR function independently to repress translation initiation. Both Hfq and art200 appear to target a putative translation enhancer located just upstream of the SD sequence, and subsequent work has found that this sequence is important for 30S ribosome binding to tnpA in vitro and translation in vivo (M. Ellis and R. Trussler, unpublished data). In this regard, translation of tnpA is repressed by intrinsic- (mRNA secondary structure and art200) and host-factors (Hfq) that cooperate to inhibit translation initiation by masking the SD sequence (secondary structure) and a translational enhancer (Hfq and art200) (Figure 6.1B, ii).

It is important to note that in the case of IS10, Hfq enhances antisense control of transposase expression but is not absolutely required for the function of RNA-OUT ((Ross et al, 2010); Chapter 2; data not shown). In this regard, Hfq acts to enhance an intrinsic regulator of transposase expression. In contrast, Hfq alone represses translation initiation on RNA-IN and tnpA and this form of regulation complements antisense control in both systems. For IS10, the dual role of Hfq in regulating transposase expression might be important for the normal biology of IS10/Tn10 elements (see section 6.2).
6.1.3 Biogenesis of a trans-acting sRNA from a transposase mRNA

The work presented in Chapter 4 shows that IS200 transposase translation is repressed by three seemingly redundant mechanisms. Secondary structure in the 5’UTR of \( \text{tnpA} \) is the strongest repressor of translation followed by art200 and a smaller contribution by Hfq. The fact that art200 binds Hfq \textit{in vivo} but Hfq is not required for antisense control was unexpected, and this prompted me to initially ask if art200 has a role independent of regulating transposition. Chapter 5 describes my work to address this question and the unexpected finding that the transposase mRNA and not art200 controls gene expression in \textit{Salmonella} Typhimurium. I initially believed that art200 could act as an Hfq-dependent sRNA in part because of an assumption that \( \text{tnpA} \) was transcribed at very low levels, like most transposase mRNAs (Nagy & Chandler, 2004). However, initial RT-qPCR experiments measuring \( \text{tnpA} \) expression, combined with Northern blot experiments, showed that \( \text{tnpA} \) was expressed at a level higher than expected and formed several stable low molecular weight species. Upon re-examining my RNA-Seq data, it became clear that \( \text{tnpA} \) overexpression had a much greater impact on \textit{Salmonella} Typhimurium gene expression than art200 depletion. The use of progressively shorter \( \text{tnpA}_{\text{trunc}} \) constructs was a key experiment for convincing us that \( \text{tnpA} \) was in fact the regulatory RNA, and this revealed that the sequence determinants for repressing SPI-1 gene expression were in the first 50 nt of \( \text{tnpA} \). Our characterization of RNA processing of the \( \text{tnpA} \) 5’UTR indicates that IS200 encodes a small noncoding RNA that is liberated from the \( \text{tnpA} \) mRNA by RNA processing. At this point I have not determined exactly what form of \( \text{tnpA} \) is active for riboregulation but for invF, it seems most likely that \( \text{tnpA}-110 \) is the functional sRNA while \( \text{tnpA}-90 \) is an inactivated form. The finding that \( \text{tnpA} \) is processed to produce \( \text{tnpA}-110 \) and then \( \text{tnpA}-90 \) revealed an additional layer of regulation for suppressing TnpA synthesis. Unlike translational control, RNA processing would irreversibly silence TnpA expression. The location of the two main processing sites implicates art200 as a key factor in directing RNA cleavage, and in this way translation repression leads to synthesis of a regulatory RNA and permanent silencing of an mRNA.
I believe that the contribution of transposon-derived sRNAs to post-transcriptional regulation of host genes is an under-explored aspect of transposon biology. In particular, regulatory RNAs could be an important way that transposons interact with the host organism. Transposon-derived sRNAs could be derived from the sense or antisense strand of transposase genes in IS elements or be encoded as a stand-alone gene in a composite transposon. For the former, there are two examples where the 3’ end of an archaeal transposase mRNA produces an sRNA (Gomes-Filho et al, 2015; Martens et al, 2013). Although my work represents the only example of a bacterial transposase gene producing an sRNA, the two examples in archaea suggest that prokaryotic transposase transcripts are a general reservoir for sRNAs. This would add transposase mRNAs to the growing list of protein-coding genes that give rise to trans-acting sRNAs (Chao & Vogel, 2016; Guo et al, 2014; Miyakoshi et al, 2015). Transposon asRNAs may also act in trans to regulate other host-encoded genes. Expression of several newly identified transposon-encoded asRNAs appears to be regulated in response to stress (Dsr30, (Tsai et al, 2015)), growth phase (art200, (Sittka et al, 2008); vc0870, (Papenfort et al, 2015)) and nitrogen availability (asRNA036, (Jager et al, 2009)). It is not obvious why an asRNA repressing transposase expression would be regulated in such a way. One simple explanation is that these asRNAs are also functioning in trans to regulate host genes. Although *tnpA* had a much greater impact on gene expression than art200, there were at least 6 genes affected by art200 depletion in *Salmonella Typhimurium*. The above asRNAs would therefore be strong candidates for transposon-encoded asRNAs that act in trans to affect host gene expression. Lastly, there is to my knowledge only a single example of an sRNA that is encoded in the non-IS sequence of a composite transposon. sRNA-Xcc1 was first identified in *Xanthomonas campestris* and homologs are found in the chromosomes or associated with integrons in a diverse range of bacteria (Chen et al, 2011) (Jiang et al, 2010). In one case, three copies of sRNA-Xcc1 are located in the plasmid-encoded Tn5542 in *Pseudomonas putida* ML2.
6.2 An integrated model for repression of IS10/Tn10 transposition by Hfq

The work presented in this thesis employed a reductionist approach to study the two distinct mechanisms by which Hfq represses Tnp synthesis in the IS10/Tn10 system. Hfq can repress translation initiation on RNA-IN directly by preventing 30S ribosome binding (Chapter 3) and indirectly by catalyzing pairing between RNA-IN and RNA-OUT (Chapter 2). However, the role of Hfq must be considered in the context of three additional features of IS10/Tn10 biology: (i) Tnp expression is repressed at the transcriptional level by Dam methylation at the RNA-IN promoter (pIN) (Case et al, 1988; Roberts et al, 1985); (ii) the strength of antisense control increases with IS10/Tn10 copy-number, in large part because RNA-OUT expression from a single element is not high enough to repress Tnp expression (Kleckner, 1990; Simons & Kleckner, 1983); (iii) only two molecules of Tnp need to be synthesized for transposition to occur (Bolland & Kleckner, 1996). The latter point is important because it means that Hfq must inhibit essentially all translation on RNA-IN to suppress transposition.

Direct repression of translation initiation on RNA-IN requires a sustained Hfq interaction to prevent 30S ribosome binding. Although Hfq is an abundant protein (~10,000 hexamers per cell) there is constant competition amongst cellular RNAs (and possibly DNA) for Hfq binding which results in continual cycling of Hfq on different RNAs (Fender et al, 2010; Hussein & Lim, 2011; Mandin & Gottesman, 2010; Moon & Gottesman, 2011; Papenfort et al, 2009; Updegrove et al, 2016). In the absence of Dam methylation, RNA-IN could be transcribed at a level high enough that the available pool of Hfq is not sufficient to completely repress Tnp synthesis (Figure 6.2A). In support of this the extent of direct repression by Hfq was inversely correlated with the amount of RNA-IN produced (cf. Figures 3.2 and 3.5). In fact, a Dam-insensitive transposase-\textit{lacZ} reporter expressed from a high-copy plasmid was almost completely insensitive to Hfq regulation (data not shown). Dam regulation of RNA-IN transcription would therefore ensure robust post-transcriptional regulation of IS10/Tn10 by Hfq (Figure 6.2B). When IS10/Tn10 is present in multi-copy the total amount of RNA-IN per cell will be higher than for a single-copy, and this would likely result in less efficient direct repression of
Tnp synthesis by Hfq. However, as IS10 copy-number increases so does the strength of antisense control by RNA-OUT. As Hfq acts catalytically to promote pairing between RNA-IN and RNA-OUT, the amount of Hfq required for total repression would be reduced (Figure 6.2C).

I envision that Hfq, Dam, and RNA-OUT cooperate throughout the IS10/Tn10 life cycle to silence transposition. If IS10 DNA is introduced to a new host as single-stranded DNA (e.g. conjugation or certain bacteriophage) there will be a short amount of time where RNA-IN transcription will be high as a consequence of unmethylated DNA. Unmethylated DNA could also be introduced from a donor cell without functioning Dam methylation. Direct repression of translation by Hfq may be impaired through competition between RNA-IN and other RNAs for Hfq-binding, and there could be a short time where IS10/Tn10 is active. For the transposon, this could be an important way to become unlinked from other mobile genetic elements (i.e. the bacteriophage or plasmid) and lead to the transposon becoming stably integrated into the genome of the new host. Once DNA methylation on pIN is established by Dam, RNA-IN transcription will be reduced and Hfq could inhibit translation completely. A similar situation could occur if IS10 DNA is introduced by natural transformation of a high-copy plasmid. The RNA-OUT promoter is only ~2.5-fold stronger than pIN, however RNA-OUT is ~50 times more stable than RNA-IN (Case et al, 1989; Simons et al, 1983). It therefore takes time for RNA-OUT to accumulate to a level high for efficient antisense control mediated by Hfq. Similar to a short amount of time without DNA methylation, this delay in establishing antisense control could provide IS10/Tn10 the time to stably integrate into the new host genome. Once RNA-OUT levels increase Hfq acts catalytically through RNA-OUT to silence Tnp synthesis.
Figure 6.2. Influence of Dam methylation and transposon copy-number on Hfq-mediated repression of IS10 transposition

(A,B) A single-copy IS10 element does not produce enough RNA-OUT (blue) for antisense control to be a significant regulator of transposase expression. Here, Hfq (green circles) represses translation on RNA-IN (red) by direct binding to the SD and blocking ribosome access. Dam-mediated methylation on pIN (B) maintains RNA-IN transcription at a level low enough that the limited pool of Hfq in the cell is able to fully repress translation. In the absence of methylation (A), RNA-IN transcription is increased to a point that competition with other mRNAs (orange) or sRNAs (light blue) for Hfq binding
results in some RNA-IN transcripts being translated. (C) The cellular concentration of RNA-OUT increases with IS10 copy-number and antisense control becomes an important regulator of transposase expression. Hfq acts catalytically to promote RNA-IN/RNA-OUT pairing and therefore less Hfq is required to fully repress transposase expression.
6.3 Is post-transcriptional regulation a general mechanism employed by host bacteria to silence transposons?

In this thesis I have shown that Hfq is a negative regulator of two different transposition systems and two distinct mechanisms of regulation have been defined. Hfq also represses IS50 and IS1413 transposase expression although for IS50 Hfq is proposed to act indirectly to repress transposase transcription, and the exact mechanism for post-transcriptional repression of IS1413 transposase has not been determined (Munshaw, 2012; Ross et al, 2014). However, it remains unclear as to the scope of Hfq regulation in bacterial transposition systems. Given the work presented in this thesis, one way to identify transposons potentially regulated by Hfq would be to mine Hfq-IP RNA-Seq data sets to identify transposase transcripts and/or putative asRNAs that bind Hfq in vivo. As an example, RNA-Seq data from an Hfq-IP performed in E. coli found that the transposase transcripts for IS5 and IS186 showed strong enrichment (7- and 10-fold) in the Hfq-IP over total RNA (Bilusic et al, 2014). A similar Hfq-IP experiment in Mycobacterium smegmatis found 5 novel asRNAs to transposase transcripts that interact with Hfq in vivo (Li et al, 2013).

Other RNA-binding proteins may play a role in regulating transposons at the post-transcriptional level. In particular, recent work identified ProQ as a major RNA-binding protein in Salmonella Typhimurium with a preference for structured asRNAs (Smirnov et al, 2016). Of note, art200 was one of the most abundant RNAs recovered in a ProQ-IP, implicating this poorly characterized host protein as a regulator of IS200. We have preliminary data showing that tnpA-lacZ expression increases in a ΔproQ strain of E. coli (R. Trussler, unpublished data) and I expect that future work will demonstrate that ProQ promotes antisense pairing with tnpA. If ProQ is required for art200-tnpA pairing this would explain my difficulty in measuring pairing despite the extensive complementarity between the two RNAs. Over 17 putative asRNAs to transposase transcripts have been identified in Gram-negative and Gram-positive bacteria (Ellis & Haniford, 2016), and ProQ and/or Hfq may be required for their function as described here for IS10.
6.4 Concluding remarks

Almost 70 years have passed since McClintock’s first proposal of a dynamic genome. The entire field of modern genetics was founded on the principle that chromosomes are unchanging and therefore predictable. Of course geneticists understood that chromosomes could accumulate small mutations and underwent recombination and segregation during meiosis but in general chromosomes were thought to be static with fixed positions for genes. McClintock’s discovery of transposable elements that could move throughout a genome and alter a phenotype was so widely dismissed that she withdrew from the scientific community within ten years of publishing her landmark papers. Transposons are now appreciated as major drivers of evolution and genome plasticity. In bacteria, transposons have a profound impact on their host and this thesis provides a small glimpse into the remarkable complexity of transposable elements. As we enter the post-genomics era it is becoming clear that there is no such thing as a standard genome and transposons play a large role in generating this diversity. I also think it likely that we will continue to discover new and unexpected ways that transposons affect their host in all domains of life.

6.5 References


Case CC, Roels SM, Jensen PD, Lee J, Kleckner N, Simons RW (1989) The unusual stability of the IS10 anti-sense RNA is critical for its function and is determined by the structure of its stem-domain. The EMBO journal 8: 4297-4305


Kittle JD, Simons RW, Lee J, Kleckner N (1989) Insertion sequence IS10 anti-sense pairing initiates by an interaction between the 5′ end of the target RNA and a loop in the anti-sense RNA. *Journal of molecular biology* **210**: 561-572

Kleckner N (1990) Regulating tn10 and is10 transposition. *Genetics* **124**: 449-454


Munshaw B (2012) The Regulation of Prokaryotic Transposable Elements by the RNA Binding Protein Hfq. MSc Thesis, University of Western Ontario,


Ross JA, Trussler RS, Black MD, McLellan CR, Haniford DB (2014) Tn5 transposition in Escherichia coli is repressed by Hfq and activated by over-expression of the small non-coding RNA SgrS. *Mob DNA* **5**: 27


Siguier P, Gourbeyre E, Chandler M (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS microbiology reviews* **38**: 865-891


Appendices

Appendix A. Copyright Permissions.

Chapter 2

Permission is granted from the RNA Society to use the following published article as part of my thesis for academic purposes without written permission as rights retained by the article authors (see http://rnajournal.cshlp.org/site/misc/permissions_landing.xhtml):


Chapter 3

Permission is granted from John Wiley and Sons to use the following published article as part of my thesis for academic purposes without written permission under the terms of the Creative Commons Attribution License (CC-BY).


Chapter 4

Written permission is provided below from Oxford University Press to use the following published article as part of my thesis for academic purposes:

Figure 1.3

Written permission is provided below from Elsevier to use Figure 1 from the following published article as part of my thesis for academic purposes:


Figure 1.4

Written permission is provided below from Macmillan Publishes Ltd to use Figure 2 from the following published article as part of my thesis for academic purposes:

OXFORD UNIVERITY PRESS LICENSE TERMS AND CONDITIONS

Oct 03, 2016

This Agreement between Michael J Ellis ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number: 3961370707293
License date: Oct 03, 2016
Licensed content publisher: Oxford University Press
Licensed content publication: Nucleic Acids Research
Licensed content title: A cis-encoded sRNA, Hfq and mRNA secondary structure act independently to suppress IS200 transposition:
Licensed content author: Michael J. Ellis, Ryan S. Trussler, David B. Haniford
Licensed content date: 27 July 2015
Type of Use: Thesis/Dissertation
Institution name: n/a
Title of your work: Riboregulation of bacterial transposons
Publisher of your work: n/a
Expected publication date: Jan 2017
Permissions cost: 0.00 CAD
Value added tax: 0.00 CAD
Total: 0.00 CAD
Requestor Location: Michael J Ellis

Attn: Michael J Ellis
Publisher Tax ID: GB125506730
Billing Type: Invoice
Billing Address: Michael J Ellis

Canada
Attn: Michael J Ellis
Total: 0.00 CAD
This Agreement between Michael J Ellis ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number: 3958830970683
License date: Sep 30, 2016
Licensed Content Publisher: Elsevier
Licensed Content Publication: Current Opinion in Microbiology
Licensed Content Title: Regulatory mechanisms employed by cis-encoded antisense RNAs
Licensed Content Author: Sabine Brantl
Licensed Content Date: April 2007
Licensed Content Volume Number: 10
Licensed Content Issue Number: 2
Licensed Content Pages: 8
Start Page: 102
End Page: 109
Type of Use: reuse in a thesis/dissertation
Portion: figures/tables/illustrations
Number of figures/tables/illustrations: 1
Format: electronic
Are you the author of this Elsevier article? No
Will you be translating? No
Order reference number
Original figure numbers: Figure 1
Title of your thesis/dissertation: Riboregulation of bacterial transposons
Expected completion date: Jan 2017
Estimated size (number of pages): 350
Elsevier VAT number: GB 494 6272 12
This Agreement between Michael J Ellis ("You") and Nature Publishing Group ("Nature Publishing Group") consists of your license details and the terms and conditions provided by Nature Publishing Group and Copyright Clearance Center.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3958841174508</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Sep 30, 2016</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
<td>Nature Publishing Group</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
<td>Nature Reviews Microbiology</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>Hfq and its constellation of RNA</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>Jörg Vogel and Ben F. Luisi</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>Aug 1, 2011</td>
</tr>
<tr>
<td>Licensed Content Volume Number</td>
<td>9</td>
</tr>
<tr>
<td>Licensed Content Issue Number</td>
<td>8</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a dissertation / thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>academic/educational</td>
</tr>
<tr>
<td>Format</td>
<td>electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>1</td>
</tr>
<tr>
<td>High-res required</td>
<td>no</td>
</tr>
<tr>
<td>Figures</td>
<td>Figure 2</td>
</tr>
<tr>
<td>Author of this NPG article</td>
<td>no</td>
</tr>
<tr>
<td>Your reference number</td>
<td></td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>Riboregulation of bacterial transposons</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Jan 2017</td>
</tr>
<tr>
<td>Estimated size (number of pages)</td>
<td>350</td>
</tr>
<tr>
<td>Requestor Location</td>
<td>Michael J Ellis</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
</tr>
</tbody>
</table>
Curriculum Vitae

MICHAEL J. ELLIS

EDUCATION

University of Western Ontario, London, ON 2011-2016
PhD Biochemistry

University of Western Ontario, London, ON 2007-2011
Honors Bachelor of Medical Sciences (BMSc), Biochemistry and Cell Biology

PUBLICATIONS


PRESENTATIONS


AWARDS

Nellie L. Farthing Memorial Fellowship in the Medical Sciences; value $3,000. University of Western Ontario (2015)

Alexander Graham Bell Canada Graduate Scholarship; value $70,000. Natural Sciences and Engineering Research Council of Canada (NSERC) (2015-2017)

Ontario Graduate Scholarship; value $15,000. Government of Ontario (2015-2016, declined)


Western Graduate Thesis Award; value $510. University of Western Ontario (2013)

USC Teaching Honor Roll. University of Western Ontario (2011-2012)

Western Graduate Research Scholarship; value $39,000. University of Western Ontario (2011-2016)

**WORK EXPERIENCE**

Instructor, Biochemistry 4410a (2016), University of Western Ontario

Teaching Assistant, Biochemistry 4410a (2013-2015), University of Western Ontario

Teaching Assistant, Biochemistry 3380g (2015), University of Western Ontario

Private Tutor (2015), London, ON

Teaching Assistant, Biochemistry 3381a (2011-2012), University of Western Ontario

Research Associate, Haniford Lab (2011), University of Western Ontario

**COMMITTEES**

Biochemistry Department Graduate Studies Committee (2014-2016), University of Western Ontario

SOGS Grad Club Committee (2013-2016), University of Western Ontario

Biochemistry Department Nominating Committee (2013-2014), University of Western Ontario