MODULATION OF EXPRESSION AND ANTIBACTERIAL TARGETING OF SHIGELLA FLEXNERI VIRF

by

Julie Kristin Hurt

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Medicinal Chemistry)
in The University of Michigan
2010

Doctoral Committee:

Associate Professor George A. Garcia, Chair
Professor Kyung-Dall Lee
Associate Professor Matthew R. Chapman
Associate Professor Anna K. Mapp
Assistant Professor Garry D. Dotson

2010

For my husband

Blake,

my parents

Doug and Karoline,

and my Kalamazoo College biology professor

Dr. Paul Olexia,

who first encouraged my career in science.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my family for their love and support over the past five years. I would especially like to thank my husband, Blake, who has weathered the storm of graduate school with me every step along the way. From celebrating my successes to consoling me through my failures, he has always provided much needed support and encouragment. I will never to be able to thank him enough for all the little, every-day thoughtfulness that has made my life so much better. I also would like to thank my parents for their support, not only through graduate school, but throughout my entire life. My father, Doug, who always encouraged me in math and science (and music), whether I cooperated or not! I am a stronger person because of these experiences, and I have achieved so much in my life and education because he made me believe it was possible. I am thankful for my mother, Karoline, who has always taught me to reach for my goals with a positive attitude and determination. She was always the first person with a word of encouragement, phone call, or greeting card to make me smile through life's difficulties. They have always been, and will continue to be, a major source of support in my life!

I would like to thank my mentor, George Garcia, for his guidance through the years. I was dead set on attending medical school when I first came to work in his lab as an undergraduate. That summer, I found an exciting research environment with a fun atmosphere that was impossible to resist. It was the best

decision I ever made! I could devote an entire chapter of this thesis to his catch-phrases (which the GAG lab has coined, “Garcia-isms”), but it’s probably better not to leave a paper trail! All joking aside, I have learned so much from him, and his advising and support has made me a better scientist. I would also like to thank my labmates, especially Dr. Jeff Kittendorf, Yi-Chen Chen and Suman Gill, who were such a big part of my experience in graduate school. I know that we have built a bond that will last a lifetime. I probably would not have made it through some months without “No Thai Tuesdays,” or our Starbucks runs! I think of you as my extended family, and your friendship has meant so much to me.

Finally, I would like to thank my committee members for their support and advice throughout the years: Professor Chapman, Professor Dotson, Professor Lee, and Professor Mapp. I would also like to thank Thomas McQuade at the Center for Chemical Genomics (Life Sciences Institute, University of Michigan) for his friendship and guidance in development of a high-throughput assay. I might not have pushed so hard to get that assay off the ground if not for his patience and advice.

This dissertation research has been supported by the National Institutes of Health (GM065489), the Pharmacological Sciences Training Program (GM077671), and the Fred W. Lyons fellowship.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

MODULATION OF EXPRESSION AND ANTIBACTERIAL TARGETING OF SHIGELLA FLEXNERI VIRF

by

Julie Kristin Hurt

Chair: George A. Garcia

Shigella flexneri is a human enteropathogen that invades the intestinal mucosa and results in severe bacillary dysentery. With the emergence of multiantibiotic-resistant strains, the development of novel therapeutics against critical *S. flexneri* target proteins promises a more effective treatment regimen for shigellosis. VirF is the positive regulator of transcription in the *Shigella spp.* virulence cascade, and Δ*virF* mutant strains of *S. flexneri* are avirulent. Although many cellular factors are associated with expression of the VirF protein (*i.e.,* temperature, pH), previous studies demonstrated that the concentration of VirF protein is decreased by 60% in Δ*vacC* mutant strains of *S.flexneri*. VacC is tRNA-guanine transglycosylase (TGT), an RNA-modification enzyme that catalyzes the incorporation of the modified nucleoside queuosine (Q) in substrate

tRNA. We have studied the VirF protein as a novel antibacterial target in *Shigella flexneri*, with a focus on both modulation of VirF expression by TGT and expression and activity of the VirF protein itself.

We hypothesized that TGT modulates VirF protein expression through queuine modification at both the level of tRNA and the *virF* mRNA. We report that the modification state of tRNA with queuine and the Q-cognate codon usage (NAU versus NAC codons) in the target mRNA alters the rate of protein expression. *VirF* mRNA contains an overall bias of 80% for the NAU Q-cognate codons, and it is conceivable that the subtle decrease in VirF protein expression in the *S. flexneri* (Δ*vacC*) mutant results from the absence of queuine-modified tRNA. In addition, we report that the *virF* mRNA is itself a substrate for the eubacterial TGT *in vitro*. Modification of RNA sequences (termed riboswitches) with small molecule metabolites has been shown to modulate either transcription or translation of the target RNA.

To measure VirF activity in the presence of small molecules, we report the development of a high-throughput cell-based reporter assay using a *virB* operator-β-galactosidase fusion. We have screened approximately 42,000 small molecules and report the identification of 7 compounds with low to mid-micromolar IC_{50} values in the VirF – β-galactosidase assay. Further studies are pending to identify selective VirF inhibitors.

CHAPTER I

Introduction

The prevalence of infectious diseases represents a continuing threat that disproportionately plagues many third world countries [1]. Responsible for one-third of all deaths, infectious diseases claim the lives of 15 million people each year, including 2 million children. Among those diseases considered to be most preventable are the diarrheal diseases, which are the cause of 15-34% of all deaths. The prevalence of such diseases is based on several unregulated components of the environment in third world countries, including access to clean drinking water, hygiene, and sanitation. Infections of the gastrointestinal tract are commonly caused by food- and water-borne pathogens (*i.e.,* enteroinvasive *E. coli*, *Shigella*, *Salmonella*, *Listeria*, etc.). These pathogens have developed complex regulatory mechanisms to infect host cells and initiate an immune response that triggers severe inflammation and cell death [2, 3]. Current treatments for bacillary dysentery include re-hydration therapy and development of a multivalent vaccine containing antigens from the following bacterial species: *Campylobacter*, *Shigella* and *Enterotoxigenic Escherichia coli* [1]. With the recent emergence of many antibiotic-resistant strains of these food- and water-borne pathogens, there is an increasing need for identification of novel agents as effective treatments for dysentery.

Shigella flexneri and Shigellosis

Shigella flexneri is a severe enteropathogen that plagues more than one million victims per year worldwide, leading to severe dysentery in humans [1]. The complex mechanism by which *Shigella* infects the host cells of the gastrointestinal tract has been the subject of much research, and was recently reviewed [2]. In the initial stages of infection, *Shigella* cross the epithelial layer in the gastrointestinal tract by way of M cells. The initial host inflammatory cascade is induced following invasion and subsequent lysis of macrophages, which release cytokines and elicit an immune response. The bacteria then invade the mucosal cells of the epithelium from the basolateral side following expression of the type III secretion system (TTSS), a bacterial virulence determinant that facilitates transfer of key virulence proteins (*i.e.,* ipa proteins) from the pathogen into the host cells [4, 5].

The pathogenicity of *Shigella* is controlled in large part by approximately 70 virulence genes encoded on a large molecular weight plasmid [6, 7]. The so-called entry region contains 37 A+T-rich open reading frames, and the protein products of these key virulence genes serve as the structural components of the TTSS as well as key regulator proteins (*i.e.,* VirB) [8]. All other virulence genes are located elsewhere on the ~220 kb plasmid [9]. There are two proteins associated with global regulation of the virulence genes: the positive regulator, VirF, and the negative regulator, H-NS (VirR) [10-13]**.** Figure I-1 presents a model of pathogenesis for *Shigella flexneri*, including several of the critical virulence proteins.

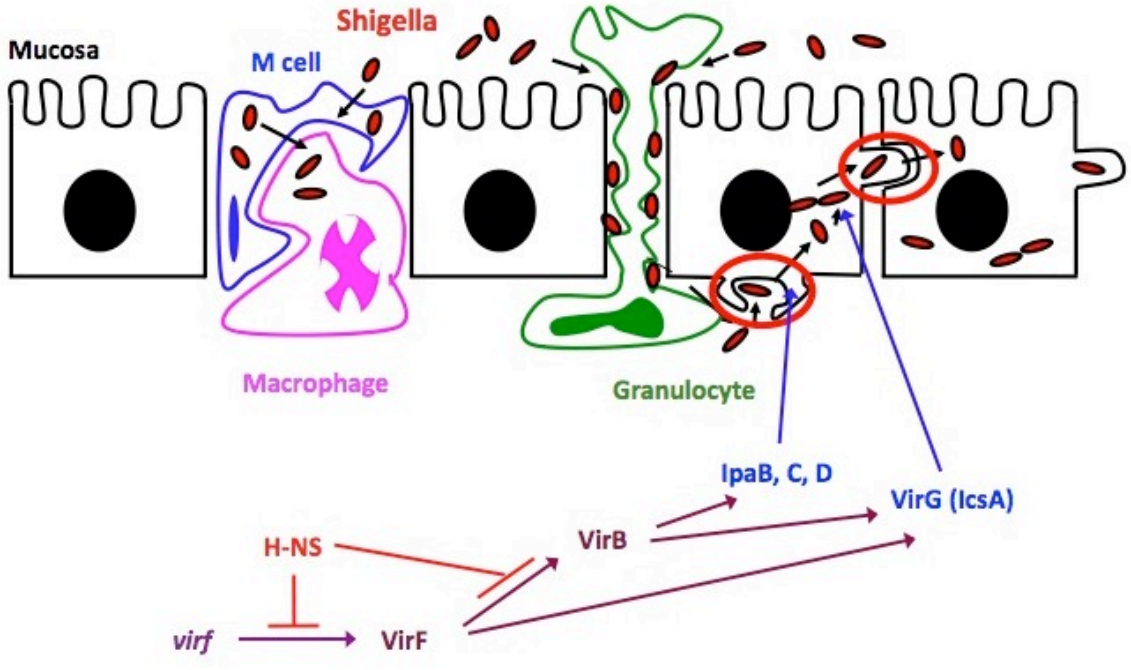


Figure I-1: Model of *Shigella* Pathogenesis. Infection of the intestinal mucosal cells is facilitated by the transcriptional activator, VirF. The secondary activator, VirB, is positively regulated by VirF and negatively regulated by H-NS (which also represses *virf* transcription). Several of the key downstream virulence proteins are also highlighted.

VirF is an AraC-type transcriptional activator that directly regulates transcription of the secondary positive regulator, VirB, and the actin-polymerizing enzyme, VirG (IcsA) [14-19]. H-NS, on the other hand, is a bacterial DNA-binding protein that forms an oligomeric structure (homodimer or -tetramer) that binds to the promoter region of many virulence genes, including *virf* and *virb*, which alters the DNA structure to a transcriptionally inactive confirmation [20, 21]. Ultimately, VirB activates transcription of many of the key downstream virulence genes (*i.e., ipaBCD, virG(icsA), mxiD*) [22, 23]. As the master regulator of positive transcriptional activation, VirF is of particular interest as a novel target in the treatment of shigellosis.

Regulation of VirF Expression

Expression of VirF in *Shigella flexneri* is tightly regulated. There are several environmental factors that promote the translation of VirF, including oxygen and iron limitation, temperature regulation, and posttranscriptional RNA modification [24]. Transcriptional repression of the *virF* and *virB* mRNAs by H-NS is regulated by temperature [25]. Full virulence gene expression is observed at 37°C, but at 30°C the H-NS oligomer binds the promoter of each gene and inhibits transcription. In the case of *virB*, there is a conformational change at 37°C that results in dissociation of H-NS and allows VirF to bind and recruit the transcriptional machinery [26]. In addition to repression by H-NS, transcription of the *virf* gene is positively regulated in a pH-dependant manner by the two-component regulatory system CpxA/CpxR [11, 27].

The modification state of several tRNAs have also been shown to play a role in translational regulation of the VirF protein [24]. One form of RNA modification is performed by the enzyme tRNA-guanine transglycosylase (TGT), which catalyzes the exchange of guanine for the modified nucleoside queuine in eukaryl and eubacterial $tRNA^{Asp,\ Asn,\ His,\ Tyr}$ [28]. *In vivo*, eubacterial organisms synthesize queuine from its chemical precursor, $preQ_1$, whereas eukaryl species must acquire queuine from external routes such as diet. The structures of the heterocyclic substrates of TGT are shown in Figure I-2.

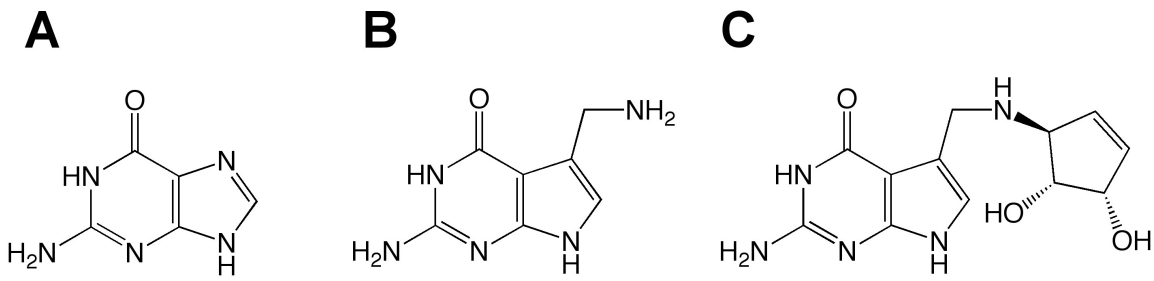


Figure I-2: Heterocyclic Substrates of TGT. A) guanine, B) $preQ_1$ (7-aminomethyl-7-deazaguanine), C) queuine ([7-(4,5-*cis*-dihydroxy-1-cyclopenten-3-yl-aminomethyl)-7-deazaguanine).

In each organism, although by different pathways, TGT catalyzes the exchange of guanine for queuine in the wobble position (34) of these cognate tRNA molecules with the common anticodon sequence $G_{34}U_{35}N_{36}$ [29]. Biochemical experiments have primarily focused on the characterization of TGT and other enzymes in the queuosine biosynthetic pathway in eubacteria [30-32]. Though much is known, and continuing to unfold about this pathway, the biological relevance of queuine-modified tRNA has not yet been fully elucidated. The *E. coli* (Δ*tgt*) mutant has been characterized as exhibiting no observable defective phenotype under optimal growth conditions, and in fact the cells grow to mid-log phase slightly faster than wildtype [33]. The *E. coli* (Δ*tgt*) mutant does exhibit growth defects under stress (particularly oxidative) conditions. Studies have also shown that while there is the potential for leaky read-through of the UAG stop codon by $tRNA^{Tyr}$-G_{34}, Q-modified $tRNA^{Tyr}$ does not mistranslate this codon [34]. Interestingly, the extent of queuine modification in eukaryotic tRNA populations has also been linked to several cancers [35, 36].

Durand and colleagues [24, 37] have demonstrated a correlation between *virF* and *tgt* by characterizing *vacC* (*tgt*) mutant *S. flexneri* strains. In these

mutants, the production of VirF is markedly reduced (VirF expression is ~40% relative to concentration in wildtype *Shigella*) and the bacteria are less virulent. When monitoring the production of the downstream virulence genes (*i.e., ipaBCD*), their expression was also reduced by approximately 50% in comparison to wildtype protein levels. Addition of an inducible copy of the *virf* gene *in trans* (*tac*-based plasmid) restored both VirF expression and virulence in the Δ*vacC Shigella flexneri* mutant. Taken together, these results suggest that the presence of Q_{34}-modified tRNA has a specific effect on the translation of the VirF protein. In order to probe this interaction between TGT and VirF, we propose two models of regulation for translation of the *virF* mRNA: 1) the modification state of queuine-cognate tRNA effects VirF expression, or 2) the *virF* mRNA is itself directly modified with queuine.

Investigating How Queuine-cognate Codon Usage Alters the Rate of Protein Expression (Chapter II).

The role modified nucleosides play in the recognition of transfer RNA at the ribosome has been an issue of ongoing investigation. Several *in vitro* studies have been conducted to monitor binding of tRNA to the ribosome-mRNA complex [38, 39]. Uhlenbeck and coworkers found that association for various tRNA species was similar, but the rate of dissociation from the ribosome was increased for unmodified tRNA versus their fully-modified counterparts [38]. While the rate of dissociation was faster overall for unmodified tRNAs, the rates varied depending on position in the ribosome (i.e., A-site versus P-site) and the state of aminoacylation. Of the Q-cognate tRNAs tested, $tRNA^{His}$ demonstrated

the most notable sensitivity to the presence of global modifications, with a dissociation rate 60-fold higher for unmodified tRNA in the P-site. Conversely, the k_{off} for $tRNA^{His}$ was only 3-fold higher for unmodified tRNA in the A-site when compared to the fully modified $tRNA^{His}$. The authors conclude that there are several unique interactions between each tRNA and the ribosome-mRNA complex, and the overall effect is to equalize the thermodynamics of each binding event to achieve a normalized rate of protein synthesis.

To determine the specific codon bias of Q-modified substrate tRNA (*i.e.,* NAU versus NAC), a cell-based approach was performed using *X. laevis* oocytes [40]. Under competitive conditions, $tRNA^{His}$-G_{34} and $tRNA^{His}$-Q_{34} from *Drosophila melanogaster* were esterified with radiolabeled histidine, and the translation of His codons in a model mRNA sequence was studied following cyanogen bromide cleavage of the protein product. When $tRNA^{His}$-G_{34} was labeled, the authors observed that peptide fragments translated from the CAC histidine codon demonstrated higher levels of radioactivity in comparison to the CAU-translated fragments (64:36, respectively). Radiolabeled $tRNA^{His}$-Q_{34}, however, did not show a significant preference for one codon over the other in comparison to the control experiment.

This apparent preference for NAC codons by Q-cognate tRNA-G_{34} was also studied computationally [41]. Molecular modeling experiments were conducted based on the crystal structure of $tRNA^{Asp}$ and the recognition of the modified/unmodified anticodon with both GAU- and GAC-containing mRNA. Queuine and several Q-precursors (guanine, 7-deazaguanine, $preQ_1$) were

modeled in position 34 of the tRNA anticodon loop. For both guanine and 7-deazaguanine, the interaction with the GAC codon was facilitated through three full Watson-Crick hydrogen bonds, while only two H-bonds were predicted with the GAU codon. The presence of $preQ_1$ and queuine in the anticodon loop altered the hydrogen bonding capability of the purine carbonyl moiety (O6) due to an energetically favorable intramolecular interaction with the amine side chain in each modified nucleoside. Interestingly, queuine was further predicted to alter the flexibility of the anticodon through an increased hydrogen bonding network, with contacts between the cyclopentendiol hydroxyl groups and the backbone of U33, as well as an intraloop hydrogen bond between U33 and C36 (Figure I-3). This interaction between U33 and C36 was first observed in molecular modeling experiments with $tRNA^{Asp}$-G_{34}, but the lifetime was shorter in the absence of queuine or $preQ_1$ [42, 43]. The authors postulate that this increased rigidity aligns the anticodon loop of the Q-modified tRNA such that only two hydrogen bonds are formed with both NAU and NAC codons, whereas tRNA-G_{34} makes preferential hydrogen bonding contact (*i.e.,* Watson-Crick hydrogen bonding) with only the NAC codon.

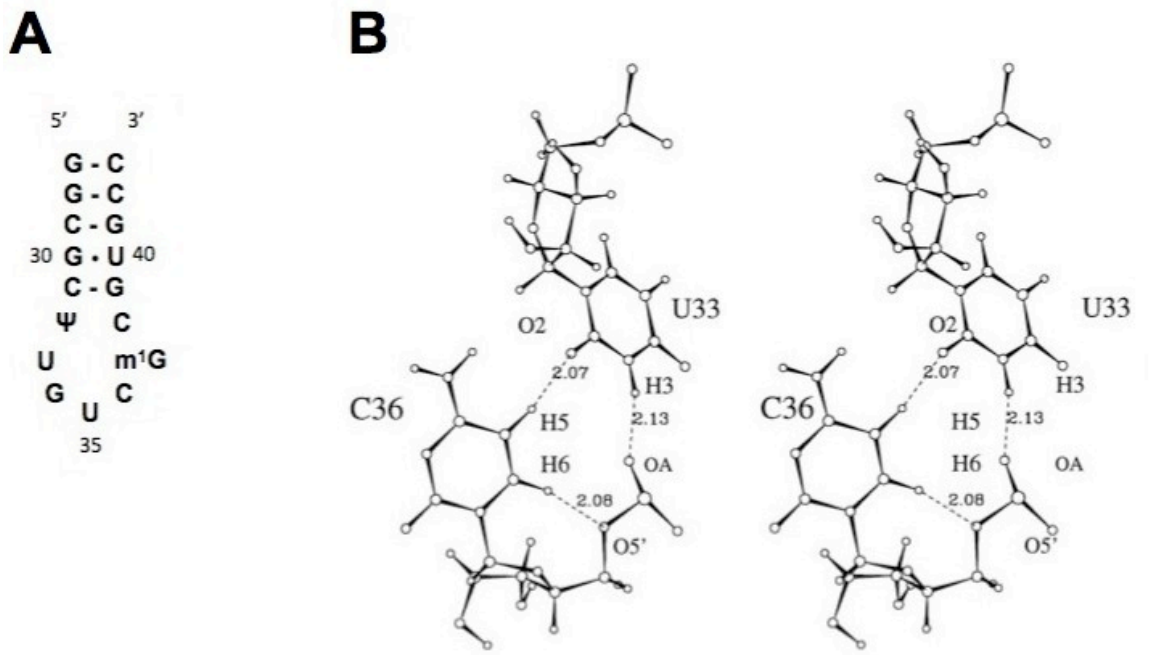


Figure I-3: Hydrogen-bonding Interactions in the Anticodon Loop of $tRNA^{Asp}$. A) Anticodon stem-loop sequence of $tRNA^{Asp}$-G_{34} (nucleotides 27-43 illustrated). B) Hydrogen-bonding parameters determined by x-ray crystallography of the $tRNA^{Asp}$ anticodon loop between uridine 33 (U33) and cytosine 36 (C36). From Auffinger *et al.* (1996), *Journal of the American Chemical Society*, **118**(5): 1181-1189.

Recently, a manuscript was published that sought to address the issue of codon bias *in vivo* using *gfp* as a model mRNA sequence [44]. A library was constructed of randomly altered gfp sequences, where the global degenerate codon usage of each template was changed by mutation of the third nucleotide position in each codon. The authors observed a significant variation in fluorescence among the constructs, suggesting the possibility that changes in codon usage directly affected the levels of GFP protein expressed. However, their primary conclusions implicate mRNA secondary structure and stability as the more influential factors on protein synthesis. Though the authors conclude that tRNA populations (i.e. rare tRNAs/isoforms) did not have a dramatic effect

on protein expression, it has been shown, in the case of *S. flexneri* VirF, that a subtle decrease in protein expression (40% in comparison to wild type levels) was sufficient to decrease activation of the virulence cascade [24]. In Chapter II, studies to probe the differential codon recognition model are presented.

Investigating Modification of virF mRNA by TGT (Chapter III)

The role of modified nucleosides in RNA structure and stability has been well-studied [45-47]. Agris and coworkers have identified key interactions between modified nucleosides and magnesium ions essential to the secondary structure of tRNA, in addition to facilitating key RNA-protein interactions [48]. Mandal and Breaker have characterized binding of small molecules and metabolites to RNA motifs termed “riboswitches” [49, 50].

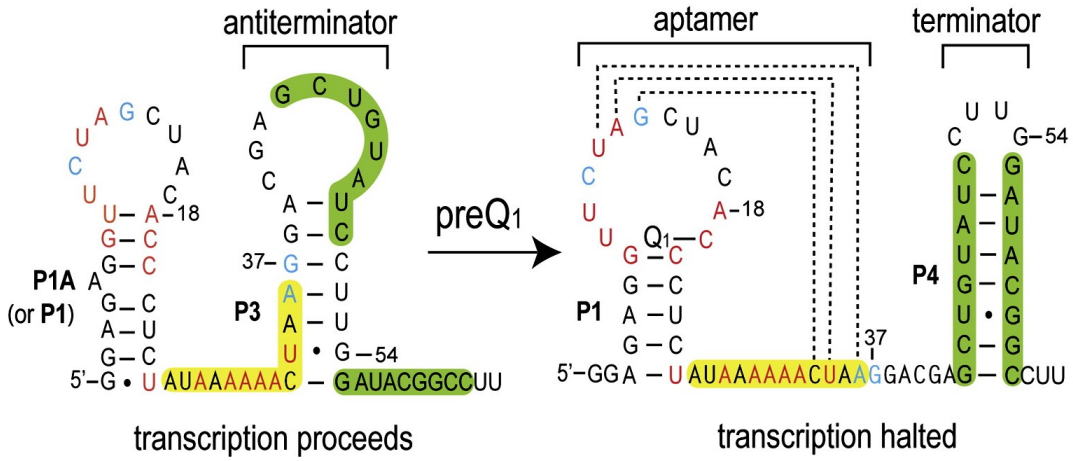


Figure I-4: Structure of the Class I PreQ$_1$ Riboswitch in *B. subtilus queC*. In the absence of preQ$_1$, the *queC* RNA forms a transcriptionally active hairpin confirmation. When preQ$_1$ is bound, however, the RNA structure changes to a terminator sequence. Taken from Kang *et al.*, *Molecular Cell* (2009), **33**(6): 784-790.

Riboswitches are structural motifs in mRNAs (often in the 5' untranslated region and extending into the start of the open reading frame) that can exist in at least

two stable conformations. One of these conformations is stabilized by binding to a small molecule, thus altering the equilibrium between the conformations. One conformation supports translation of the protein while with the other conformation, translation is blocked. In this way, binding of the small molecule changes the conformation of the RNA and modulates its translation. Riboswitches have also been characterized as regulators of transcription, with the formation of a transcription terminator complex dependent on the concentration of the bound metabolite [51]. In addition to the 5' untranslated region (UTR) of prokaryotic mRNAs, riboswitches have also been found in the 3' UTR and introns in several eukaryotic species [52]. Interestingly, several groups have reported the discovery of two classes of riboswitches that respond to the queuine precursor, $preQ_1$, and appear to regulate the expression of the four genes that are involved in $preQ_1$ biosynthesis [53-56]. In the case of a $preQ_1$ riboswitch identified in the queuosine biosynthetic gene, *queC*, solution NMR structures revealed that $preQ_1$-bound RNA formed a transcription terminator pseudoknot structure (Figure I-4) [56]. It certainly seems possible that base modification of an mRNA could modulate a similar conformational change. It therefore is feasible that such a control mechanism for gene expression might be involved in regulating the expression of virulence factors in pathogenic organisms, as is apparently seen with VirF.

The incorporation of modified nucleosides has been characterized more fully for tRNA (and some other RNAs, *e.g.,* rRNA & snRNA), than for mRNA [57-59]. The most common example of post-transcriptional processing in mRNA is

the eukaryotic 7-methylguanosine 5' cap structure, which aids in the binding to the small ribosomal subunit and is essential for the efficient synthesis of eukaryotic proteins [60-62]. To date, the only known function of TGT is to catalyze the modification of tRNA with queuine. Previous work has shown that the eubacterial TGT will recognize a U-G-U sequence in the loop of an RNA hairpin structure that corresponds to the anticodon stem-loop of its cognate tRNAs [63, 64]. The eubacterial TGT will also recognize a U-G-U containing hairpin in the context of a dimeric form of a cognate tRNA [65]. It is conceivable that an mRNA may also be modified directly by TGT provided the mRNA contained the appropriate recognition elements. In Chapter III, *in vitro* kinetic analyses of *E. coli* TGT and various *virF* mRNA substrates will be discussed.

The VirF Protein – A Novel Drug Target in *Shigella*

In addition to monitoring the regulation of VirF expression, our second aim is to study VirF function. VirF is a member of the AraC family of transcriptional regulators [66]. There are three classes of AraC-type regulators: chemically-modulated, physically-modulated, and monomeric [67]. AraC itself is a member of the first class and is modulated by arabinose [68]. VirF is a member of the second class and is modulated by temperature, pH and osmolarity [69]. MarA is a member of the third class of monomeric regulators that lack the amino terminal dimerization domain [70, 71]. In general, a high degree of sequence similarity is exhibited in the C-terminal domain of AraC-family members, but very little sequence homology is observed in the N-terminus (where applicable). Figure I-5

shows a 1.6 Å crystal structure solved for AraC in the L-arabinose-bound and unbound states [68].

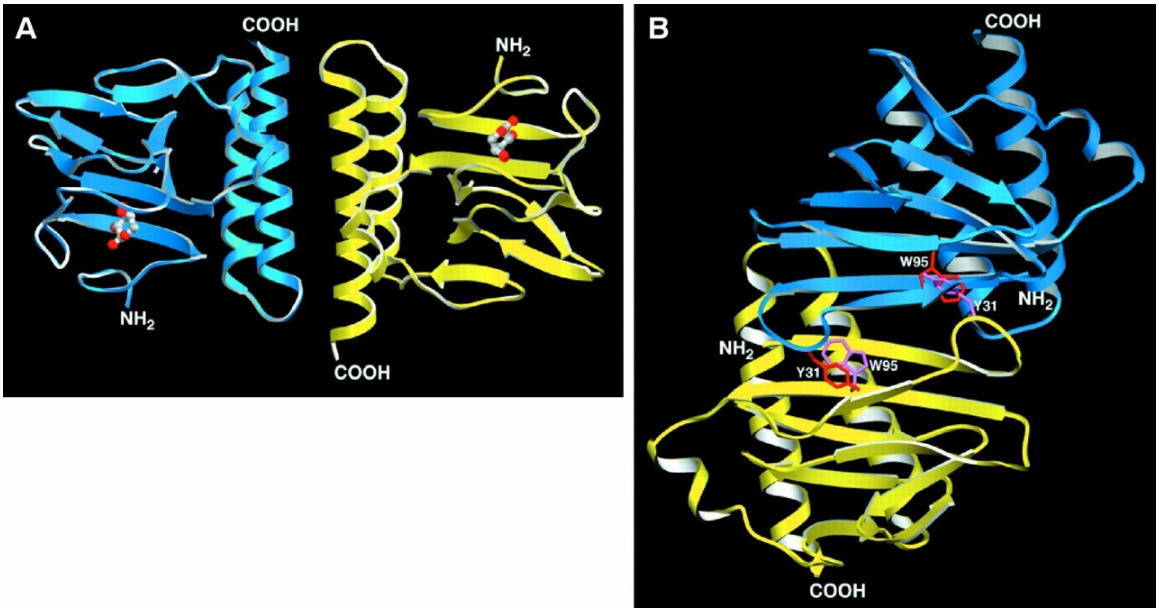


Figure I-5: Crystal Structure of AraC from *E. coli*. A) In the presence of L-arabinose, the DNA-binding domains of the AraC monomers are held in close proximity for binding of adjacent sites in the downstream promoters. B) In the absence of arabinose, the DNA-binding domains are bound to opposing regions of the promoter DNA, and transcription initiation is halted. Taken from Soisson *et al. Science*, (1997), **276**(5311): 421-425.

In most cases, AraC-family members function as transcriptional activators; however, AraC is unique in that the homodimer acts as both a repressor and an activator of the ara_{BAD} operon. In the absence of L-arabinose, AraC binds opposing sites in the *araBAD* operon and creates a DNA loop that prevents transcription from occurring [72, 73]. In the presence of L-arabinose, however, the AraC dimer binds an adjacent site in the operon, releasing the DNA loop and adopting a conformation favorable for transcription initiation.

Many bacterial pathogens utilize AraC-type regulators to activate transcription of key virulence genes (*i.e., Shigella* (VirF), *Enterotoxigenic E. coli* (CfaD),

Salmonella (SprA), *Vibrio* (ToxT), *Pseudomonas* (ExsA)) [69, 74-77]. The relatively widespread distribution of VirF and VirF-homologous systems suggests that the investigation of VirF as a novel antibiotic target in *Shigella* may extend to other important human pathogens. However, many attempts to express and purify recombinant AraC-family members have failed due to the insoluble nature of the overexpressed DNA-binding proteins [16, 78]. Due to increased solubility, the monomeric class of transcriptional activators has been studied more extensively [70, 71].

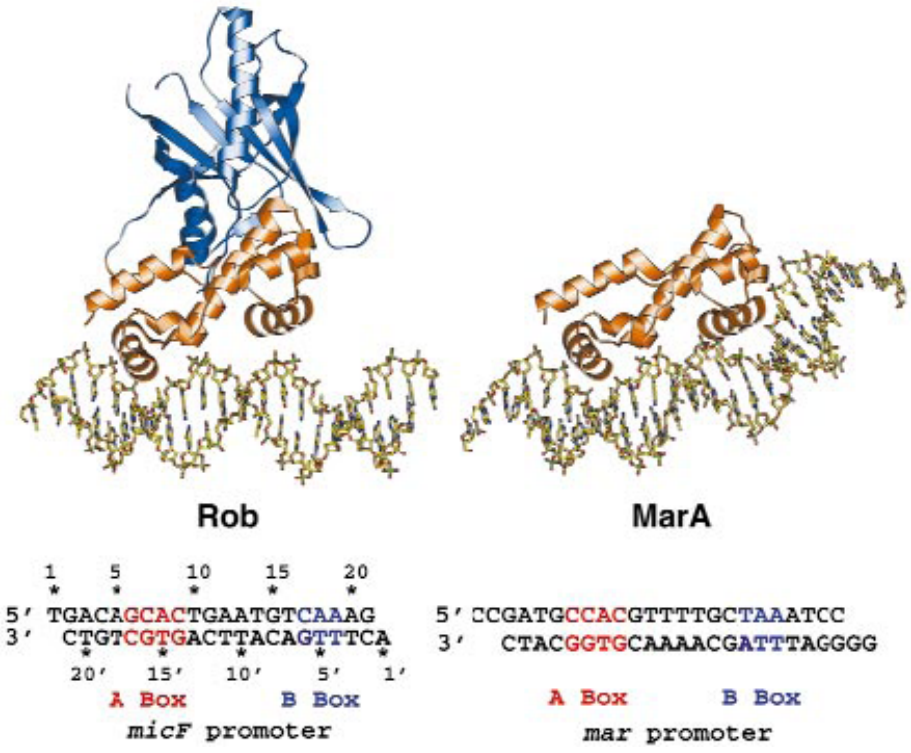


Figure I-6: Crystal Structures of Monomeric AraC-type Proteins Rob and MarA with DNA Bound. Rob binds to the micF promoter making primary contacts with one of the HTH motifs. MarA (bound to the mar promoter) facilitates DNA-binding with both HTH motifs characteristic of the AraC family, creating a bend in the DNA. Taken from Kwon *et al., Nature Struct. Biol.* (2000), **7**, 424-430.

Figure I-6 shows the crystal structures of Rob and MarA bound to their DNA promoter sequences (micF and mar respectively), with the critical protein-DNA interactions made through the conserved helix-turn-helix motif (HTH) characteristic to the AraC family. Although much information has been obtained from DNA-binding analyses of this monomeric class, many of the AraC-type regulators involved in bacterial virulence function as homodimers of multi-domain proteins (*i.e.,* contain both dimerization and DNA-binding domains). In an attempt to gain more knowledge of the virulence factor, VirF from *S. flexneri*, we propose 1) study of the recombinant VirF protein *in vitro* with various molecular biology manipulations in an attempt to improve expression, and 2) expression in a cell-based reporter assay as a high-throughput screen for potential inhibitors.

Expression and Purification Trials with the VirF Protein (Chapter IV)

Discovery of new methods for expressing “problematic” or “toxic” proteins *in vitro* and *in vivo* is an ongoing area of research. In some cases, expression is studied through alteration of the protein induction conditions in a host strain (typically *E. coli*). One standard method for increasing protein solubility is the expression of recombinant proteins with fusion tags and has been the subject of several recent reviews [79-81]. Fusion tags are selected based on optimal characteristics of protein solubility and ease of expression. Among the most commonly used fusion tags are MBP (maltose-binding protein), GST (glutathione-S-transferase), Trx (thioredoxin) and NusA (N-utilization substance) [79]. Although in some cases these protein partners can result in improved solubility, the effects of each fusion tag are different depending on the candidate

protein of interest. Other methods for altering the induction conditions in *E. coli* include lower growth temperatures and over expression of specific tRNAs for expression of proteins with rare codon usage [82].

Protein insolubility is often linked to aggregation of the recombinant protein in *E. coli* and the formation of inclusion bodies [83]. Some evidence suggests that expression of insoluble proteins can be overcome with use of more tightly regulated bacterial promoters, such as the araBAD promoter from *E. coli* [84]. Molecular chaperones have also been used to direct protein folding when co-expressed with the target protein. Several heat shock proteins with chaperone function have been identified in prokaryotic species (*i.e.,* GroEL, GroES, DnaK, GrpE) [85, 86]. If solubility cannot be regulated under native expression conditions, some success has come from denaturing the protein of interest in the presence of high concentrations of a chemical denaturant (*i.e.,* urea or guanidinium chloride). Protein re-folding is then conducted with gradual dialysis to remove the denaturant and exchange the protein into a proper buffer to monitor recovery of activity. Though much is known about troubleshooting this issue of protein solubility, expression of the recombinant VirF protein in *E. coli* has historically proven difficult [16, 87]. In Chapter IV, a presentation of efforts to express recombinant VirF *in vivo* and *in vitro* will be discussed.

Development of a High-throughput Reporter Assay to Screen Small Molecule Inhibitors of VirF Activity (Chapter V)

Recently, there has been increased interest in development of novel antibiotics that target virulence factors [88-90]. Though the prevalence of AraC-

type transcriptional activators make them an attractive target for treatment of a wide range of bacterial pathogens, the pharmaceutical industry has historically focused on development of bactericidal antimicrobial agents (*i.e.,* inhibitors of DNA/RNA, protein, or cell wall synthesis). Though many potent drugs have been designed against these fundamental bacterial targets, the use of antibiotics in the food industry and everyday household products has lead to the emergence of antibiotic-resistant strains of medically-relevant bacteria [91]. Because expression of virulence factors is not required for cell viability, there should be less selective pressure for the pathogens to develop resistance to inhibitors of such targets. Though the study of AraC family members has thus far focused on identification of downstream virulence targets and protein-DNA interactions, there is increasing interest in developing functional assays to screen for potential inhibitors.

Researchers at Paratek Pharmaceuticals have studied several members of the AraC family related to bacterial virulence *in vitro* [92, 93]. Using a coupled approach with *in silico* screening of small molecules and solved crystal structures of three AraC members from *E. coli* (MarA, SoxS and Rob) and an *in vitro* DNA-binding assay, researchers identified a promising class of AraC-type DNA-binding inhibitors: hydroxybenzimidazole derivatives [93]. Though the compounds demonstrated inhibition of DNA-binding *in vitro* with IC_{50} values in the low micromolar range, this study only addresses one aspect of AraC-type regulator function. Recently, the same group reported activity of a similar compound set (*N*-hydroxbenzimidazoles) against LcrF in *Yersinia spp.* [92]. LcrF

is a Mar-like protein that activates transcription of the type III secretion system. Following synthesis of various analogues, compounds were validated in a cell-based cytotoxicity assay as well as DNA-binding analysis with an LcrF homologue (ExsA from *Salmonella*) and an unrelated transcriptional activator (SlyA from *Salmonella*) to establish specificity. In the case of VirF, there are at least three steps in the gene activation process (although the order is not clearly understood): DNA binding, dimerization, and recruitment of the transcription complex. In Chapter V, we describe a primary cell-based functional assay, which would screen for inhibition of any of these processes, with more specific secondary assays to both validate the hits and to probe their mechanisms of action.

Research Objectives

VirF from *Shigella flexneri* is the key transcriptional activator of the virulence cascade. Though some details of VirF activity and specificity are known (*i.e.,* DNA-binding sites, primary targets for activation, activators and repressors of protein expression), the specific mechanism by which VirF activates transcription is unclear. One interesting point of regulation in the VirF lifecycle was discovered by Durand and colleagues, demonstrating a link between RNA modification by TGT and protein expression [24, 94]. To better understand this master virulence regulator, the approach for this dissertation research is two-fold: 1) characterize the mechanism by which RNA modification with queuine alters VirF protein expression, and 2) study the VirF protein directly to gain structural and functional information about the bacterial transcription

factor. The primary questions addressed in each chapter of this dissertation are listed below.

RNA Modification and VirF Expression

- Chapter II: Does the queuosine modification in substrate tRNA and degenerate codon usage alter the rate of protein expression?
- Chapter III: Does the *virF* mRNA serve as a substrate for queuosine modification by TGT?

Functional/Structural Characterization of the VirF Protein

- Chapter IV: Can the issue of recombinant VirF solubility in *E. coli* be overcome by molecular manipulation?
- Chapter V: Can we study VirF *in vivo* as a means to screen small molecule inhibitors of the bacterial transcription factor?

Notes to Chapter I

1. WHO, *State of the Art of New Vaccines: Research & Development*. 2003, Initiative for Vaccine Research, World Health Organization: Geneva.

2. Dorman, C.J., *The Virulence Plasmids of Shigella flexneri*, in *Microbial Megaplasmids*. 2009, Springer: Berlin / Heidelberg. p. 151-170.

3. Dramsi, S. and P. Cossart, *Intracellular pathogens and the actin cytoskeleton.* Annual Review of Cell and Developmental Biology, 1998. **14**: p. 137-166.

4. Allaoui, A., P.J. Sansonetti, and C. Parsot, *MxiD, an outer-membrane protein necessary for the secretion of the Shigella flexneri ipa invasins.* Molecular Microbiology, 1993. **7**(1): p. 59-68.

5. Kenjale, R., et al., *The needle component of the type III secreton of Shigella regulates the activity of the secretion apparatus.* Journal of Biological Chemistry, 2005. **280**(52): p. 42929-42937.

6. Sansonetti, P.J., D.J. Kopecko, and S.B. Formal, *Involvement of a plasmid in the invasive ability of Shigella flexneri.* Infection and Immunity, 1982. **35**(3): p. 852-860.

7. Sasakawa, C., et al., *Molecular alteration of the 140-megadalton plasmid associated with loss of virulence and congo red binding-activity in Shigella flexneri.* Infection and Immunity, 1986. **51**(2): p. 470-475.

8. Buchrieser, C., et al., *The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri.* Molecular Microbiology, 2000. **38**(4): p. 760-771.

9. Hale, T.L., *Genetic basis of virulence in Shigella species.* Microbiological Reviews, 1991. **55**(2): p. 206-24.

10. Berlutti, F., et al., *Expression of the virulence plasmid-carried apyrase gene (apy) of enteroinvasive Escherichia coli and Shigella flexneri is under the control of H-NS and the VirF and VirB regulatory cascade.* Infection & Immunity, 1998. **66**(10): p. 4957-64.

11. Nakayama, S. and H. Watanabe, *Identification of cpxR as a positive regulator essential for expression of the Shigella sonnei virF gene.* Journal of Bacteriology, 1998. **180**(14): p. 3522-8.

12. Prosseda, G., et al., *The virF promoter in Shigella: more than just a curved DNA stretch.* Molecular Microbiology, 2004. **51**(2): p. 523-37.

13. Schuch, R. and A.T. Maurelli, *Virulence plasmid instability in Shigella flexneri 2a is induced by virulence gene expression.* Infection & Immunity, 1997. **65**(9): p. 3686-92.

14. Adler, B., et al., *A dual transcriptional activation system for the 230 kb plasmid genes coding for virulence-associated antigens of Shigella flexneri.* Molecular Microbiology, 1989. **3**(5): p. 627-35.

15. Jost, B.H. and B. Adler, *Site of transcriptional activation of virB on the large plasmid of Shigella flexneri 2a by VirF, a member of the AraC family of transcriptional activators.* Microbial Pathogenesis, 1993. **14**(6): p. 481-8.

16. Tobe, T., et al., *Transcriptional control of the invasion regulatory gene virB of Shigella flexneri: activation by virF and repression by H-NS.* Journal of Bacteriology, 1993. **175**(19): p. 6142-9.

17. Lett, M.C., et al., *VirG, a plasmid-coded virulence gene of Shigella flexneri - Identification of the VirG protein and determination of the complete coding sequence.* Journal of Bacteriology, 1989. **171**(1): p. 353-359.

18. Suzuki, T., S. Saga, and C. Sasakawa, *Functional analysis of Shigella VirG domains essential for interaction with vinculin and actin-based motility.* Journal of Biological Chemistry, 1996. **271**(36): p. 21878-21885.

19. Wing, H.J., et al., *Regulation of IcsP, the outer membrane protease of the Shigella actin tail assembly protein IcsA, by virulence plasmid regulators VirF and VirB.* Journal of Bacteriology, 2004. **186**(3): p. 699-705.

20. Porter, M.E. and C.J. Dorman, *A role for H-NS in the thermo-osmotic regulation of virulence gene expression in Shigella flexneri.* Journal of Bacteriology, 1994. **176**(13): p. 4187-4191.

21. Stella, S., et al., *Environmental control of the in vivo oligomerization of nucleoid protein H-NS.* Journal of Molecular Biology, 2006. **355**(2): p. 169-174.

22. Dorman, C.J., *Virulence gene regulation in Shigella*, in *Escherichia coli and Salmonella: cellular and molecular biology*, I.R. Curtiss, et al., Editors. 2004, American Society for Microbiology: Washington, D.C.

23. Turner, E.C. and C.J. Dorman, *H-NS antagonism in Shigella flexneri by VirB, a virulence gene transcription regulator that is closely related to plasmid partition factors.* Journal of Bacteriology, 2007. **189**(9): p. 3403-3413.

24. Durand, J.M., et al., *Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of Shigella flexneri: the expression of the virF gene.* Molecular Microbiology, 2000. **35**(4): p. 924-35.

25. Maurelli, A.T., B. Blackmon, and R. Curtiss, *Temperature-dependent expression of virulence genes in Shigella species.* Infection and Immunity, 1984. **43**(1): p. 195-201.

26. Tobe, T., et al., *Temperature-regulated expression of invasion genes in Shigella flexneri is controlled through the transcriptional activation of the virB gene on the large plasmid.* Molecular Microbiology, 1991. **5**(4): p. 887-893.

27. Nakayama, S. and H. Watanabe, *Involvement of CpxA, a sensor of a 2-component regulatory system, in the pH-dependent regulation of expression of Shigella sonnei virF gene.* Journal of Bacteriology, 1995. **177**(17): p. 5062-5069.

28. Harada, F. and S. Nishimura, *Possible Anticodon Sequences of $tRNA^{His}$, $tRNA^{Asn}$, and $tRNA^{Asp}$ from Escherichia coli B. Universal Presence of Nucleoside Q in the First Position of the Anticodons of These Transfer Ribonucleic Acids.* Biochemistry, 1972. **11**: p. 301-308.

29. Tsang, T.H., M. Buck, and B.N. Ames, *Sequence Specificity of tRNA-Modifying Enzymes.* Biochimica Biophysica Acta, 1983. **741**: p. 180-196.

30. Reader, J.S., et al., *Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine.* Journal of Biological Chemistry, 2004. **279**(8): p. 6280-6285.

31. Grosjean, H., V. de Crecy-Lagard, and G.R. Bjork, *Aminoacylation of the anticodon stem by a tRNA-synthetase paralog: relic of an ancient code?* Trends in Biochemical Sciences, 2004. **29**(10): p. 519-522.

32. Iwata-Reuyl, D., *Biosynthesis of the 7-deazaguanosine hypermodified nucleosides of transfer RNA.* Bioorganic Chemistry, 2003. **31**(1): p. 24-43.

33. Noguchi, S., et al., *Isolation and Characterization of an Escherichia coli Mutant Lacking tRNA-Guanine Transglycosylase.* Journal of Biological Chemistry, 1982. **257**(11): p. 6544-6550.

34. Bienz, M. and E. Kubli, *Wild-type $tRNA^{Tyr}$ reads the TMV RNA stop codon, but Q base-modified $tRNA^{Tyr}$ does not.* Nature, 1981. **294**: p. 188-190.

35. Pathak, C. and M. Vinayak, *Modulation of lactate dehydrogenase isozymes by modified base queuine.* Molecular Biology Reports, 2005. **32**(3): p. 191-196.

36. Pathak, C., Y.K. Jaiswal, and M. Vinayak, *Queuine mediated inhibition in phosphorylation of tyrosine phosphoproteins in cancer.* Molecular Biology Reports, 2008. **35**(3): p. 369-374.

37. Durand, J.M. and G.R. Bjork, *Putrescine or a combination of methionine and arginine restores virulence gene expression in a tRNA modification-deficient mutant of Shigella flexneri: a possible role in adaptation of virulence.* Molecular Microbiology, 2003. **47**(2): p. 519-27.

38. Fahlman, R.P., T. Dale, and O.C. Uhlenbeck, *Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites.* Molecular Cell, 2004. **16**(5): p. 799-805.

39. Olejniczak, M., et al., *Idiosyncratic tuning of tRNAs to achieve uniform ribosome binding.* Nature Structural & Molecular Biology, 2005. **12**(9): p. 788-93.

40. Meier, F., et al., *Queuosine modification of the wobble base in tRNA^{His} influences 'in vivo' decoding properties.* EMBO J., 1985. **4**: p. 823-827.

41. Morris, R.C., K.G. Brown, and M.S. Elliott, *The effect of queuosine on tRNA structure and function.* Journal of Biomolecular Structure & Dynamics, 1999. **16**(4): p. 757-774.

42. Auffinger, P. and E. Westhof, *H-bond stability in the tRNA(Asp) anticodon hairpin: 3 ns of multiple molecular dynamics simulations.* Biophysical Journal, 1996. **71**(2): p. 940-954.

43. Auffinger, P. and E. Westhof, *RNA hydration: Three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin.* J Mol Biol, 1997. **269**(3): p. 326-341.

44. Kudla, G., et al., *Coding-Sequence Determinants of Gene Expression in Escherichia coli.* Science, 2009. **324**(5924): p. 255-258.

45. Agris, P.F., *The importance of being modified: Roles of modified nucleosides and Mg2+ in RNA structure and function*, in *Progress in Nucleic Acid Research and Molecular Biology, Vol 53*, W.E. Cohn and K. Moldave, Editors. 1996, Academic Press Inc: 525 B Street, Suite 1900, San Diego, CA 92101-4495. p. 79-129.

46. Varani, G. and I. Tinoco, *RNA Structure and NMR-Spectroscopy.* Quarterly Reviews Of Biophysics, 1991. **24**(4): p. 479.

47. Heus, H.A. and A. Pardi, *Structural Features That Give Rise To The Unusual Stability Of Rna Hairpins Containing Gnra Loops.* Science, 1991. **253**(5016): p. 191.

48. Agris, P.F. and S.C. Brown, *Systems for the NMR study of modified nucleoside-dependent, metal-ion induced conformational changes in nucleic acids*, in *Nuclear Magnetic Resonance and Nucleic Acids*, T.L. James, Editor. 1995, Academic Press Inc: 525 B Street, Suite 1900, San Diego, CA 92101-4495. p. 270-299.

49. Barrick, J.E., et al., *New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control.* Proceedings Of The National Academy Of Sciences Of The United States Of America, 2004. **101**(17): p. 6421.

50. Mandal, M. and R.R. Breaker, *Gene regulation by riboswitches.* Nature Reviews Molecular Cell Biology, 2004. **5**(6): p. 451.

51. Coppins, R.L., K.B. Hall, and E.A. Groisman, *The intricate world of riboswitches.* Current Opinion in Microbiology, 2007. **10**(2): p. 176-181.

52. Sudarsan, N., J.E. Barrick, and R.R. Breaker, *Metabolite-binding RNA domains are present in the genes of eukaryotes.* Rna-a Publication of the Rna Society, 2003. **9**(6): p. 644-647.

53. Meyer, M.M., et al., *Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria.* RNA-A PUBLICATION OF THE RNA SOCIETY, 2008. **14**(4): p. 685-95.

54. Roth, A., et al., *A riboswitch selective for the queuosine precursor preQ(1) contains an unusually small aptamer domain.* Nature Structural & Molecular Biology, 2007. **14**(4): p. 308-317.

55. Klein, D.J., T.E. Edwards, and A.R. Ferre-D'Amare, *Cocrystal structure of a class I preQ(1) riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase.* Nature Structural & Molecular Biology, 2009. **16**(3): p. 343-344.

56. Kang, M., R. Peterson, and J. Feigon, *Structural Insights into Riboswitch Control of the Biosynthesis of Queuosine, a Modified Nucleotide Found in the Anticodon of tRNA.* Molecular Cell, 2009. **33**(6): p. 784-790.

57. Grosjean, H., M. Sprinzl, and S. Steinberg, *Posttranscriptionally modified nucleosides in transfer RNA: Their locations and frequencies.* Biochimie, 1995. **77**(1-2): p. 139-141.

58. Gu, X.G., J. Ofengand, and D.V. Santi, *In Vitro Methylation of Escherichia coli 16S rRNA by tRNA (m^5U54)-Methyltransferase.* Biochemistry, 1994. **33**(8): p. 2255-2261.

59. Noon, K.R., E. Bruenger, and J.A. McCloskey, *Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile Sulfolobus solfataricus.* J Bacteriol, 1998. **180**(11): p. 2883-2888.

60. Pain, V.M., *Initiation of protein synthesis in eukaryotic cells.* European Journal Of Biochemistry, 1996. **236**(3): p. 747.

61. Macejak, D.G. and P. Sarnow, *Internal Initiation Of Translation Mediated By The 5' Leader Of A Cellular Messenger RNA.* Nature, 1991. **353**(6339): p. 90.

62. Gingras, A.C., B. Raught, and N. Sonenberg, *eIF4 initiation factors: Effectors of mRNA recruitment to ribosomes and regulators of translation.* Annual Review Of Biochemistry, 1999. **68**: p. 913.

63. Curnow, A.W., et al., *tRNA-Guanine Transglycosylase from Escherichia coli: Gross tRNA Structural Requirements for Recognition.* Biochemistry, 1993. **32**: p. 5239-5246.

64. Curnow, A.W. and G.A. Garcia, *tRNA-Guanine Transglycosylase from Escherichia coli - Minimal tRNA Structure and Sequence Requirements for Recognition.* Journal of Biological Chemistry, 1995. **270**(29): p. 17264-17267.

65. Curnow, A.W. and G.A. Garcia, *tRNA-Guanine Transglycosylase from Escherichia coli: Recognition of Dimeric, Unmodified tRNA^{Tyr}.* Biochimie, 1994. **76**(12): p. 1183-1191.

66. Dorman, C.J. and M.E. Porter, *The Shigella virulence gene regulatory cascade: a paradigm of bacterial gene control mechanisms.* Molecular Microbiology, 1998. **29**(3): p. 677-684.

67. Martin, R.G. and J.L. Rosner, *The AraC transcriptional activators.* Current Opinion in Microbiology, 2001. **4**(2): p. 132-137.

68. Soisson, S.M., et al., *Structural basis for ligand-regulated oligomerization of AraC.* Science, 1997. **276**(5311): p. 421-425.

69. Tobe, T., M. Yoshikawa, and C. Sasakawa, *Thermoregulation of virB transcription in Shigella flexneri by sensing of changes in local DNA superhelicity.* Journal of Bacteriology, 1995. **177**(4): p. 1094-7.

70. Rhee, S., et al., *A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(18): p. 10413-8.

71. Kwon, H.J., et al., *Crystal structure of the Escherichia coli Rob transcription factor in complex with DNA.* Nature Structural Biology, 2000. **7**(5): p. 424-30.

72. Timmes, A., M. Rodgers, and R. Schleif, *Biochemical and physiological properties of the DNA binding domain of AraC protein.* Journal of Molecular Biology, 2004. **340**(4): p. 731-8.

73. Schleif, R., *Regulation of the L-arabinose operon of Escherichia coli.* Trends in Genetics, 2000. **16**(12): p. 559-65.

74. Eichelberg, K., W.D. Hardt, and J.E. Galan, *Characterization of SprA, an AraC-like transcriptional regulator encoded within the Salmonella typhimurium pathogenicity island 1.* Molecular Microbiology, 1999. **33**(1): p. 139-152.

75. Pilonieta, M.C., M.D. Bodero, and G.P. Munson, *CfaD-dependent expression of a novel extracytoplasmic protein from enterotoxigenic Escherichia coli.* Journal of Bacteriology, 2007. **189**(14): p. 5060-5067.

76. Hsiao, A., et al., *Direct Regulation by the Vibrio cholerae Regulator ToxT To Modulate Colonization and Anticolonization Pilus Expression.* Infection and Immunity, 2009. **77**(4): p. 1383-1388.

77. Brutinel, E.D., et al., *Characterization of ExsA and of ExsA-dependent promoters required for expression of the Pseudomonas aeruginosa type III secretion system.* Molecular Microbiology, 2008. **68**(3): p. 657-671.

78. Schleif, R., *AraC protein: a love-hate relationship.* Bioessays, 2003. **25**(3): p. 274-82.

79. Esposito, D. and D.K. Chatterjee, *Enhancement of soluble protein expression through the use of fusion tags.* Current Opinion in Biotechnology, 2006. **17**(4): p. 353-358.

80. Sorensen, H.P. and K.K. Mortensen, *Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli.* Microbial Cell Factories, 2005. **4**: p. 8.

81. Li, Y.F., *Carrier proteins for fusion expression of antimicrobial peptides in Escherichia coli.* Biotechnology and Applied Biochemistry, 2009. **54**: p. 1-9.

82. Kleber-Janke, T. and W.M. Becker, *Use of modified BL21(DE3) Escherichia coli cells for high-level expression of recombinant peanut allergens affected by poor codon usage.* Protein Expression & Purification, 2000. **19**(3): p. 419-424.

83. de Groot, N.S., et al., *Studies on bacterial inclusion bodies.* Future Microbiology, 2008. **3**(4): p. 423-435.

84. Martinez-Alonso, M., et al., *Learning about protein solubility from bacterial inclusion bodies.* Microbial Cell Factories, 2009. **8**: p. 5.

85. Lin, Z. and H.S. Rye, *GroEL-Mediated protein folding: Making the impossible, possible.* Critical Reviews in Biochemistry and Molecular Biology, 2006. **41**(4): p. 211-239.

86. Muga, A. and F. Moro, *Thermal Adaptation of Heat Shock Proteins.* Current Protein & Peptide Science, 2008. **9**(6): p. 552-566.

87. Porter, M.E. and C.J. Dorman, *In vivo DNA-binding and oligomerization properties of the Shigella flexneri AraC-like transcriptional regulator VirF as identified by random and site-specific mutagenesis.* Journal of Bacteriology, 2002. **184**(2): p. 531-9.

88. Alekshun, M.N. and S.B. Levy, *Targeting virulence to prevent infection: to kill or not to kill?* Drug Discovery Today: Therapeutic Strategies, 2004. **1**(4): p. 483-489.

89. Barczak, A.K. and D.T. Hung, *Productive steps toward an antimicrobial targeting virulence.* Current Opinion in Microbiology, 2009. **12**(5): 490-496.

90. Clatworthy, A.E., E. Pierson, and D.T. Hung, *Targeting virulence: a new paradigm for antimicrobial therapy.* Nature Chemical Biology, 2007. **3**(9): p. 541-548.

91. DuPont, H.L. *The growing threat of foodborne bacterial enteropathogens of animal origin*. in *15th Annual James H Steels DVM Meeting*. 2007. Houston, TX: Univ Chicago Press.

92. Kim, O.K., et al., *N-hydroxybenzimidazole inhibitors of the transcription factor LcrF in Yersinia: novel antivirulence agents.* J Med Chem, 2009. **52**(18): p. 5626-34.

93. Bowser, T.E., et al., *Novel anti-infection agents: small-molecule inhibitors of bacterial transcription factors.* Bioorganic & Medicinal Chemistry Letters, 2007. **17**(20): p. 5652-5.

94. Durand, J.M., et al., *vacC, a Virulence-associated Chromosomal Locus of Shigella flexneri, is Homologous to tgt, a Gene Encoding tRNA-Guanine Transglycosylase (TGT) of Escherichia coli K-12.* Journal of Bacteriology, 1994. **176**(15): p. 4627-4634.

CHAPTER II

tRNA Wobble Base Modification Modulates Translation via a Redundant Codon Bias Mechanism

The genetic code provides multiple synonymous codons for 18 of the 20 naturally occurring amino acids (excluding methionine and tryptophan). For many of these amino acids, there are several tRNA isoforms that recognize a particular codon or subset of codons. Codon usage varies between species, and the implications of codon bias were recently reviewed [1]. Many factors are involved in protein synthesis, including the concentration of appropriate tRNA isoforms, the post-transcriptional modification state of tRNA and the energetics of tRNA binding to the mRNA-ribosome complex [2, 3]. Though the occurrence of degenerate codon usage has been well documented, there is little direct evidence measuring the biological relevance of this phenomenon.

Recently, a study was published that sought to address the effect of codon bias on protein translation *in vivo* using *gfp* as a model mRNA [4]. The author's primary conclusions implicate mRNA secondary structure and stability as the more influential factors on protein synthesis. Although these authors observed that tRNA populations (*i.e.,* rare tRNAs/isoforms) did not have a dramatic effect on protein expression, it has been shown that the modification state of tRNA can

affect not only protein synthesis but bacterial virulence as well in the case of a Δ*vacC* (*tgt* homologue) mutant of *Shigella flexneri* [5]. In this mutant strain, a subtle decrease in protein expression of the positive transcriptional regulator, VirF (40% in comparison to wild type levels) was sufficient to decrease activation of the virulence cascade.

tRNA-guanine transglycosylase (TGT) is an RNA modification enzyme that catalyzes the incorporation of the non-cannonical nucleoside, queuosine (Q), into the wobble position of four substrate tRNAs (Asn, Asp, His, Tyr) [6]. For each of these four amino acids, there is a single tRNA that decodes two synonymous codons (NAC and NAU, where N is any of the canonical nucleosides). Interestingly, the *virF* mRNA exhibits a bias of ≥90% for NAU codons of Asp, His, and Tyr, with an overall bias of 80% toward the NAU Q-cognate codons. In this study, we aim to address the biological relevance of the queuine modification in substrate tRNA as it relates to protein synthesis. With DNA microarray analysis comparing the gene expression profiles of wild-type *E. coli* K12 and a K12(Δ*tgt*) mutant, we report several key cellular mRNAs with altered expression levels, relative to wild-type, in the absence of this unique RNA modification. Computational analysis of the protein-coding sequences in the genomes of *E. coli* and *S. flexneri* revealed similar trends in protein classes with biased queuine-cognate codon usage. Finally, we show with detailed kinetic analyses that both tRNA-G_{34} and tRNA-Q_{34} demonstrate apparent preferential recognition of the degenerate Q-cognate codons in expression of GFP.

Materials and Methods

Reagents

Unless otherwise specified, all reagents were from Sigma or Aldrich. DNA oligonucleotides, agarose, dithiothreitol (DTT), T4 DNA ligase, IPTG, SuperScriptIII™ reverse transcriptase, RNaseOUT RNase inhibitor, and DNA ladders were from Invitrogen. All restriction enzymes and Vent® DNA polymerase were from New England Biolabs. The deoxyribonucleic acid triphosphates (dNTPs) were from Promega. Low-melting Seaplaque agarose was from Cambrex. Gelase™ Enzyme Prep, ScriptGuard™ RNase inhibitor, and MasterPure™ RNA Purification Kit were from Epicentre. BugBuster® Protein Extraction Reagent, Lysonase™ Bioprocessing Reagent, and pET21b were from Novagen. Epicurian coli® XL2-Blue ultracompetent cells were from Stratagene. Gene Pulser electroporation cuvettes were purchased from BioRad. Amicon Ultra Centrifugal Filter Devices were from Millipore. Whatman GF/C Glass Microfibre Filters and carbenicillin were from Fisher Scientific. The QIAPrep® Spin Miniprep Kit was from Qiagen. The GeneChip Ecoli_ASv2 and data analysis program were from Affymetrix. Synthetic *gfp* plasmid constructs were prepared by GenScript. The CybrGreen master mix, real-time PCR reagents and data analysis software were from Applied Biosystems.

DNA Microarray Analysis

E. coli K12 was from ATCC and K12(Δ*tgt*) was received as a gift from Dr. Klaus Reuter (University of Marburg, Germany) [7]. Liquid cultures (100 mL 2xTY) were inoculated with starter culture for each cell line (1:100) and grown at

37°C with shaking. Aliquots (1 mL) were removed every 30 minutes from each sample to monitor bacterial growth by UV-Vis (OD_{600}). Total RNA was isolated from both K12 and K12(Δ*tgt*) at mid-log and stationary phases using the MasterPure™ RNA Purification Kit according to the vendor's protocol. Generation of cDNA and microarray experiments were performed by the Molecular Biology Core Facility (Dental School, University of Michigan). The gene expression profile in each sample was analyzed with the GeneChip Ecoli_ASv2 according to the vendor's protocol.

Computational Genomic Analysis

The complete genomes for *E. coli* K12 (substrain MG1655, NC_000913) and *S. flexneri* 2a (substrain 301, NC_004337), as well as the *Shigella* virulence plasmid, pWR501 (NC_002698), were obtained from the NCBI database. The synonymous codon usage for each protein-coding open reading frame was tabulated with the computational tool CodonW [8]. The data was analyzed for pronounced codon bias in the queuine-cognate codons (NAC and NAU) as determined by the following criteria: 1) NAC = NAC/(NAC+NAU) ≥ 0.75, or 2) NAU = NAU/(NAC+NAU) ≥ 0.8 in at least three of the four Q-cognate codons. Candidate open reading frames were cross-referenced with the NCBI database to determine the function (known or predicted) of each protein-coding sequence identified.

Determining the Rate of GFP Expression

Construction of codon-biased gfp vectors

The sequence for wild-type *gfp* DNA was obtained from the NCBI database (U50963). The *gfp* transcript consists of 239 codons, 51 of which are queuine-cognate. Synthetic *gfp* constructs corresponding to the wild-type sequence and *gfp* with engineered Q-cognate codon bias (GFP(C) and GFP(U)) were prepared by GenScript. GFP(C) contains 100% NAC codons and GFP(U) contains 100% NAU codons. Each sample was constructed with restriction enzyme sites for recognition by *Eco*RI (3') and *Hind*III (5') for cloning into pET21b. Each *gfp* open reading frame and the pET21b vector were treated in a double restriction enzyme digestion with *Eco*RI and *Hind*III (40 U each, 20 µL reaction) for 1 h at 37°C. All samples were gel-purified from Seaplaque agarose gel with Gelase™ according to the vendor's protocol. Each *gfp* insert was subcloned into pET21b (5:1 volume ratio, respectively) with T4 DNA ligase (2 U, 20 µL reaction) at 17°C overnight. The ligation products (10 µL) were transformed into chemically competent *E. coli* TG2 cells (JM101 derivative) and grown on L-Carb plates (100 µg/mL carbenicillin) at 37°C. Liquid 2xTY cultures (3 mL with carbenicillin) were inoculated with individual colonies and grown at 37°C with shaking. Plasmid DNA was isolated from each sample with the QIAprep Spin Miniprep Kit according to vendor's protocol. The resultant GFP expression vectors (pET21b-*gfp*(wt), *gfp*(C), and *gfp*(U)) were confirmed by DNA sequencing (University of Michigan DNA Sequencing Core Facilities).

Kinetic analysis of GFP expression

Electrocompetent *E. coli* K12(DE3) and K12(Δ*tgt*, DE3) cells were prepared and transformed with each pET21b-*gfp* sample according to the

MicroPulser Electroporation protocol (BioRad). Cells were grown overnight at 37°C on L-Carb plates. Individual colonies were isolated for growth in 10 mL 2xTY (with 100 µg/mL carbenicillin), and each construct was assayed in replicate (n=6). All samples were incubated at 37°C with shaking to mid-log phase (OD_{600} = 0.6-0.7). Aliquots (1 mL) were removed from each sample to a pre-tared microcentrifuge tube as an uninduced control. Samples were induced with IPTG (1 mM) and divided in half, with one set of replicates induced at 37°C with shaking and the other set at 30°C. Aliquots (1 mL) were removed at each hour of the 4 h induction period, and the cells were harvested by centrifugation. Following aspiration of the media, the pellets were weighed and subsequently resuspended in BugBuster® Protein Extraction Reagent (120 µL) supplemented with Lysonase™ Bioprocessing Reagent (1 µL) and PMSF (100 mM) for 5 min. at room temperature. Each sample (100 µL) was added to a black, flat-bottom microtiter plate (Corning), and GFP expression was monitored by fluorescence ($F_{485/518\ nm}$).

The rate of expression for GFP(wt), GFP(C), and GFP(U) in both *E. coli* K12(DE3) and K12(Δ*tgt*, DE3) was determined from a plot of $F_{(485/518\ nm)}$ / mg cell pellet versus time (h). The data were fit by linear regression using the first four data points; the 4 h time point was not used in the rate calculation because in most cases it was outside of the linear range. The average rate of protein synthesis from each Q-cognate codon biased *gfp* transcript was determined from the six replicate samples and presented with the corresponding standard deviation. The expression of each GFP construct was reported relative to

GFP(wt) in wild-type K12(DE3) at each temperature. The standard deviation is a measure of the relative error between samples calculated by the following equation: when y = a/b, then $\Delta y = ((\Delta a/a)-(\Delta b/b))\bullet y$, where "y" is the relative expression, "a" is the rate of expression in the test sample, and "b" is the rate of expression in GFP(wt) in K12(DE3).

Quantification of gfp mRNA levels

Isolation of total RNA from induced cultures

All pET21b-*gfp* constructs were assayed in both *E. coli* K12(DE3) and K12(Δ*tgt*, DE3) in duplicate according to the above protocol with some modifications. Upon induction with IPTG, the samples were incubated in parallel at 37°C and 30°C for 3 h. Aliquots (1 mL) were removed from each sample and placed in pre-tared microcentrifuge tubes, and the cells were harvested by centrifugation. The total RNA (with induced *gfp* mRNA) was extracted from the cell pellets as described previously. The final RNA pellet was resuspended in TE Buffer (25 µL) with ScriptGuard™ RNase Inhibitor (40 U). The final concentration of RNA was determined by UV-Vis using the NanoDrop spectrophotometer (Thermo Scientific).

Northern Blot (Passive Diffusion)

The method for Northern blot analysis of ^{32}P-labeled *gfp* mRNA is a modified denaturing agarose gel/passive transfer protocol developed by Katrin Karbstein and colleagues (Young and Karbstein, unpublished). The method and results are described in Appendix II-1.

Reverse Transcription

Gfp cDNA was synthesized from each duplicate total RNA preparation using a modified procedure from Invitrogen. 16S rRNA was chosen as a reasonable endogenous control for standardization of RNA content and stability. Total RNA (10 µg) was incubated with antisense primers for either *gfp* (test RNA) or 16S rRNA (endogenous control) in parallel (4 pmol each): *gfp* REV: 5'–CACTTGTACAGCTCGTCCATGCCAAG–3', and 16S REV: 5'–CATGGCTGCATCAGGCTTGCGC–3'. Briefly, reactions containing total RNA, gene-specific oligos, and dNTP mix (0.5 mM each NTP) in a total volume of 26 µL were denatured at 65°C for 5 min. and cooled immediately on ice for 1 min. The following components were added to each sample (final volume of 40 µL): first-strand buffer (250 mM Tris•HCl (pH 8.3), 375 mM KCl, 15 mM $MgCl_2$), DTT (5 mM), SuperScriptIII™ reverse transcriptase (200 U), RNaseOUT™ RNase Inhibitor (40 U), and DMSO (10%). Samples were incubated at 55°C for 1 h for efficient first-strand synthesis.

Real-time PCR

All samples were submitted for real-time PCR analysis at the Molecular Biology Core Facility (Dental School, University of Michigan). HPLC-purified oligos were purchased for amplification of *gfp* and 16S cDNA (see REV primers above): *gfp* FWD: 5'–GAGCAAGGGCGAGGAACTGTTCACTG–3', and 16S FWD: 5'–GAAGAGTTTGATCATGGCTCAG–3'. Amplification of both nucleic acids was monitored in the ABI PRISM® 7700 Sequence Detection System thermocycler with CybrGreen Master mix (Applied Biosystems). Briefly,

detection of *gfp* cDNA was monitored from samples containing reverse transcription (RT) product (1 µL) and *gfp* HPLC-purified oligos (5 pmol each) in a total volume of 30 µL. Reactions were tested in replicate (n=4) and incubated according to the following temperature sequence: 40 cycles – 95°C for 15 sec., 60°C for 30 sec., and 72°C for 30 sec. For detection of 16S rRNA, samples containing RT product (1 µL, 100-fold dilution) and 16S HPLC-purified oligos (15 pmol each) in a total volume of 30 µL. Reactions were tested in replicate (n=4) and incubated according to the following temperature sequence: 40 cycles – 95°C for 15 sec., 60°C for 30 sec., and 72°C for 45 sec.

Statistical Analysis

The average threshold cycle (C_T) for *gfp* and 16S cDNA synthesis in each sample was determined. The expression of each *gfp* codon-biased construct was measured by averaging the C_T data for duplicate biological samples, and reporting *gfp* cDNA expression relative to the levels of *gfp*(wt) expression in wild-type K12(DE3) at each corresponding temperature. The standard deviation was calculated according to the same equation described for the GFP protein analysis. Statistical significance was determined with the Student's t-test.

Results

DNA Microarray Analysis

To probe the physiological roles of tRNA-guanine transglycosylase in *E. coli*, the global gene expression profiles in both wild-type *E. coli* K12 and *E. coli* K12(Δ*tgt*) (hereafter termed K12 and K12(Δ*tgt*)) were detremined via DNA microarray analyses. A comparison of the expression patterns at both mid-log

and stationary phase revealed key differences in mRNA transcript levels that were reduced or increased in K12(Δ*tgt*) relative to wild-type K12 (Table II-1). Patterns in biological function of affected transcripts emerged. Most notably, five of the most reduced RNA transcripts in K12(Δ*tgt*) at mid-log phase belong to the *fim* operon, responsible for synthesis of type-1 fimbrial proteins. Of the ten genes with the most significant increase in mRNA concentration at mid-log phase, five encode for enzymes that prepare biologically relevant substrates for glycolysis and the citric acid cycle. In addition, several transcripts related to carbohydrate-utilizing enzymes and tRNAs were shown to have a differential expression pattern in K12(Δ*tgt*) in comparison to wild-type K12 over the growth cycle. The relationship in biological function of affected transcripts was less apparent in stationary phase, with the appearance of hypothetical and predicted proteins becoming more predominant. For these genes identified by DNA microarray analysis, no pronounced bias in the queuine-cognate codons was observed (as defined in Materials and Methods). In general, genes with decreased expression in K12(Δ*tgt*) did exhibit a bias for NAU in one or two of the cognate codons, and genes with increased expression demonstrated a bias for NAC in one or two of the cognate codons.

Genome Analysis for Queuine-Cognate Codon Usage in E. coli and S. flexneri

Although a strong codon bias was not observed for genes identified by microarray, genes necessary for their activation/function may show a pronounced bias. To probe this possibility, the protein-coding sequences of the *E. coli* K12

and *S. flexneri* 2a genomes were analyzed with the biophysical tool CodonW [8]. The algorithm in CodonW tabulates the codon usage patterns of degenerate codons for each of the twenty amino acids. The data were sorted to identify genes with a strong bias in at least three of the four queuine-cognate codons to more closely match the bias seen in *virF* (Figure II-3). The bias for NAU (NAU codons / (NAU+NAC codons)) and NAC (NAC codons / (NAU+NAC codons)) Q-cognate codons were determined for each species, and the results were grouped according to biological function as described by homology sequence alignments in the NCBI protein database (Figure II-1, full list of genes found in Appendix II-3). The threshold for pronounced bias of 80% was initially chosen because this is the observed bias for *virF*. This was maintained for the NAU-biased genes, but the threshold was set to 75% in the case of NAC-biased genes because the number of hits was much smaller at a threshold of 80%.

As seen with the DNA microarray data, interesting trends in protein functional classes emerged. Open reading frames demonstrating a strong NAU-bias in both *E. coli* and *S. flexneri* aligned with the biological functions we observed as being down-regulated in the K12(Δ*tgt*) mutant: fimbrial proteins and membrane-associated proteins (Figure II-1A and C, respectively). Also, genes involved in the biosynthesis of lipopolysaccharides were identified, which comprise key structural membrane components in bacterial species. Several components of the phosphotransferase system involved in sugar uptake (referenced in the “enzymes” functional class) were also identified as having a pronounced NAU-bias. It should be noted that of the 20 NAU-biased genes

Table II-1: DNA Microarray Global Gene Expression Profile of *E. coli* K12 and K12 (Δ*tgt*). The top results for genes with differential expression in the presence or absence of *tgt* are highlighted.

Mid-log phase: E. coli K12 (Δtgt) versus E. coli K12			*Stationary phase: E. coli K12 (Δtgt) versus E. coli K12*		
Gene ID	**Gene Description**	**Fold Change**	**Gene ID**	**Gene Description**	**Fold Change**
Genes with Decreased Expression			*Genes with Decreased Expression*		
fimI	Fimbrial protein, type-1 pilus biosynthesis	-140.10	b2656	Hypothetical protein	-180.06
fimA	Major type 1 subunit fimbrin (pilin)	-85.39	yrhA	Hypothetical protein	-171.00
glyY	Glycine tRNA3	-60.05	b1471	Hypothetical protein	-151.67
fimD	Outer membrane usher protein, type-1 fimbrial synthesis	-43.20	agaB	N-acetylgalactosamine-specific enzyme IIB component of PTS	-111.89
argX	Arginine tRNA3	-42.08	proM	Proline tRNA3	-109.98
fimC	Pilin chaperone, periplasmic	-39.93	yjgG	Hypothetical protein	-106.23
glpQ	Periplasmic glycerophosphodiester phosphodiesterase	-29.91	ycgW	Predicted protein	-104.75
fimF	Minor component of type-1 fimbriae	-28.48	yidZ	Predicted DNA-binding transcriptional regulator	-97.32
glpK	Glycerol kinase	-20.73	yedN	Hypothetical protein	-94.97
glpT	Sn-glycerol-3-phosphate transporter	-19.70	b2857	Hypothetical protein	-91.32
Genes with Increased Expression			*Genes with Increased Expression*		
ydaA	Stress-induced protein	386.11	mcrA	5-methylcytosine-specific restriction endonuclease B	28.26
betB	Betaine aldehyde dehydrogenase	126.39	ssrS	6S RNA, interacts with holo RNAP	26.37
narG	Nitrate reductase I, alpha subunit	105.86	ymfM	Hypothetical protein	24.60
aceA	Isocitrate lyase	86.95	ymfE	Predicted inner membrane protein	23.62
cstC	Succinylornithine transaminase	84.52	cynX	Predicted cyanate transporter	19.60
lacA	Thiogalactoside acetyltransferase	83.50	b1145	E14 prophage; repressor protein	19.00
b0370	Hypothetical protein	73.50	lacA	Thiogalactoside acetyltransferase	18.33
aceB	Malate synthase A	69.15	glpF	Glycerol facilitator	17.82
yahO	Predicted protein	68.82	yahM	Hypothetical protein	17.76
glcB	Malate synthase G	64.87	rpsM	30S ribosomal subunit protein S13	17.48

identified as nucleic acid-binding proteins in *E. coli*, 11 were classified as DNA-binding transcriptional regulators (same functional class as VirF). A number of predicted/hypothetical proteins were also identified as NAU-biased genes in both genomes, though the percentage is larger in *S. flexneri*. This is most likely due to the fact that the *E. coli* genome has been more carefully annotated.

As with the NAU-biased analysis, genes identified with a pronounced NAC-bias clustered into various functional classes. Nearly 30% of the genes identified as having a strong NAC-bias in both *E. coli* and *S. flexneri* are ribosomal protein subunits or other components of the translational machinery (Figure II-1B and D, respectively). One ribosomal open reading frame (*rpsM*) demonstrated increased expression in K12(Δ*tgt*) by microarray. Although more ribosomal subunits were not listed, four additional ribosomal subunits (rpsGKLN) are found among the top 20 transcripts with increased expression in K12(Δ*tgt*).

The genes encoded on the 230 kb virulence plasmid from *S. flexneri* were also analyzed for Q-cognate codon bias (Figure II-2). Only 2 genes were identified as having a pronounced bias toward NAC codons (*tnpR* and *tnpA*, related to the Tn501 transposition processes). Of the nearly 300 virulence genes, 24 were identified with bias of ≥ 80% toward NAU codons. VirF was the only transcriptional regulator to demonstrate such a strong bias. The genes identified belong predominantly to open reading frames responsible for synthesis of the type III secretion system and secreted invasion proteins (*mxi-spa* operons).

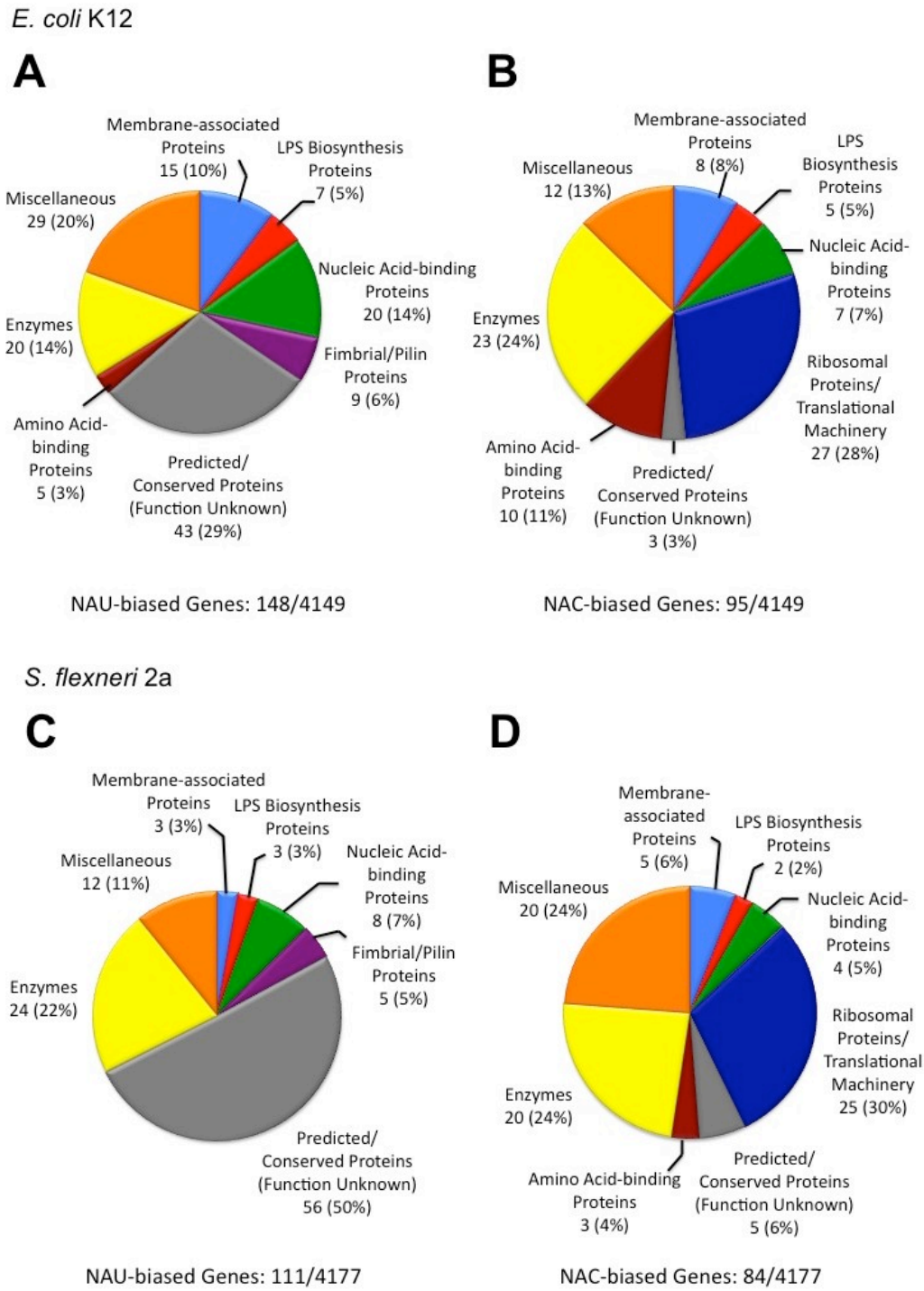


Figure II-1: Comparison of Queuine-Cognate Codon Bias in the *E. coli* and *S. flexneri* Genomes. Protein-coding sequences were analyzed for bias in at least three of the four Q-cognate codons as described in the materials and methods. Genes demonstrating bias are presented according to biological function, number of genes, and percentage of total genes.

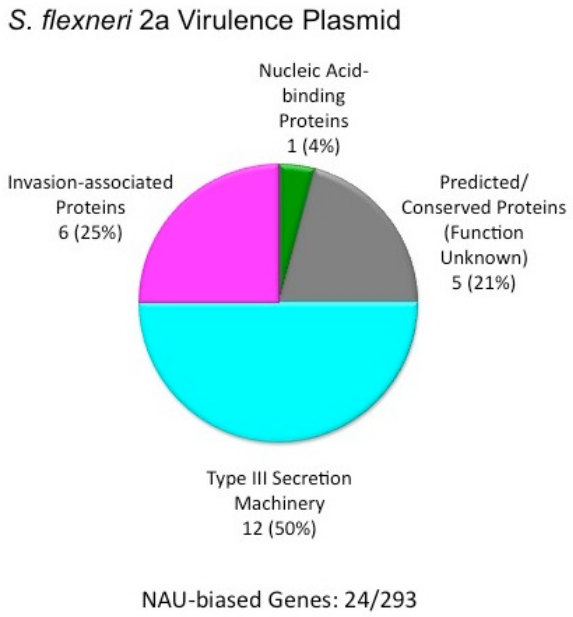


Figure II-2: Queuine-Cognate Codon Bias in the *S. flexneri* Virulence Plasmid. Protein-coding sequences were analyzed for bias in at least three of the four Q-cognate codons. The genes demonstrating bias are presented according to biological function, number of genes, and percentage of total genes.

Analysis of Protein Expression from Engineered Biased Constructs

The genes identified by DNA microarray and genomic analyses belong to functional protein classes that made experimental expression studies and assays relatively difficult (*i.e.,* membrane-associated proteins, large molecular weight protein complexes). A suitable model protein was found in green fluorescent protein (GFP), which was initially selected based on the ease of assay and the distribution and ratio of queuine-cognate codons (Figure II-3). The wild-type *gfp* mRNA contains a slight bias for NAC codons over NAU codons (63%, Table II-2). Three open reading frames for *gfp* (wild-type, 100% NAU and 100% NAC in all 51 Q-cognate codons; *gfp*(wt), *gfp*(U) and *gfp*(C), respectively) were engineered by GenScript. Each construct was expressed in K12(DE3) and K12(Δ*tgt*, DE3) at

both 30°C and 37°C. Determination of GFP concentration in cell lysates was measured by fluorescence over a 4 h incubation period and normalized for the weight of the resuspended cell pellet.

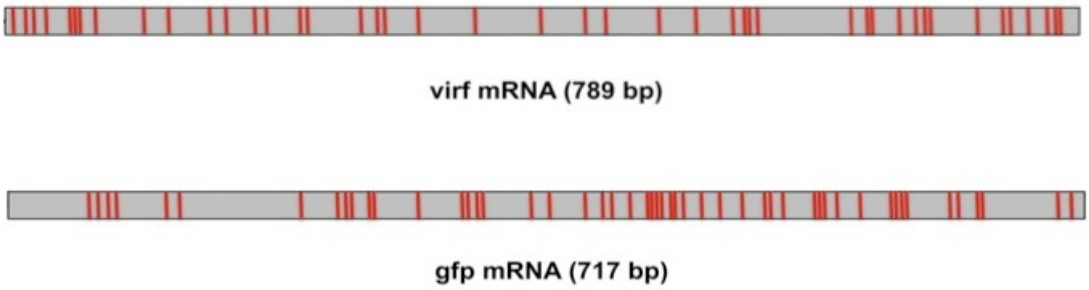


Figure II-3: Queuine-Codon Distribution in *virF* and *gfp* mRNA. Positioning of the NAU and NAC codons are highlighted in red.

Table II-2: Comparison of Codon Usage in *virF* and *gfp* mRNA.

mRNA Sequence	Total Codons	NAC Codons	NAU Codons	Clusters (≥2)
virF	263	9	35	6
gfp	239	32	19	10

Samples were assayed in replicates of 6, and the data were fit by linear regression (see Figure II-A3-A in Appendix II-2, inset for a representative set). The average rate of expression for each GFP sample is shown in Figure II-A3-A and Table II-3. The deviation observed is reflective of the variability in GFP expressed from individual colonies, and not the variability within a given sample. The expression of each *gfp* construct at 30°C demonstrated slightly less variability between independent cultures (average percent error in 37ºC samples of 19% versus 18% in 30ºC samples). The levels of codon-biased GFP

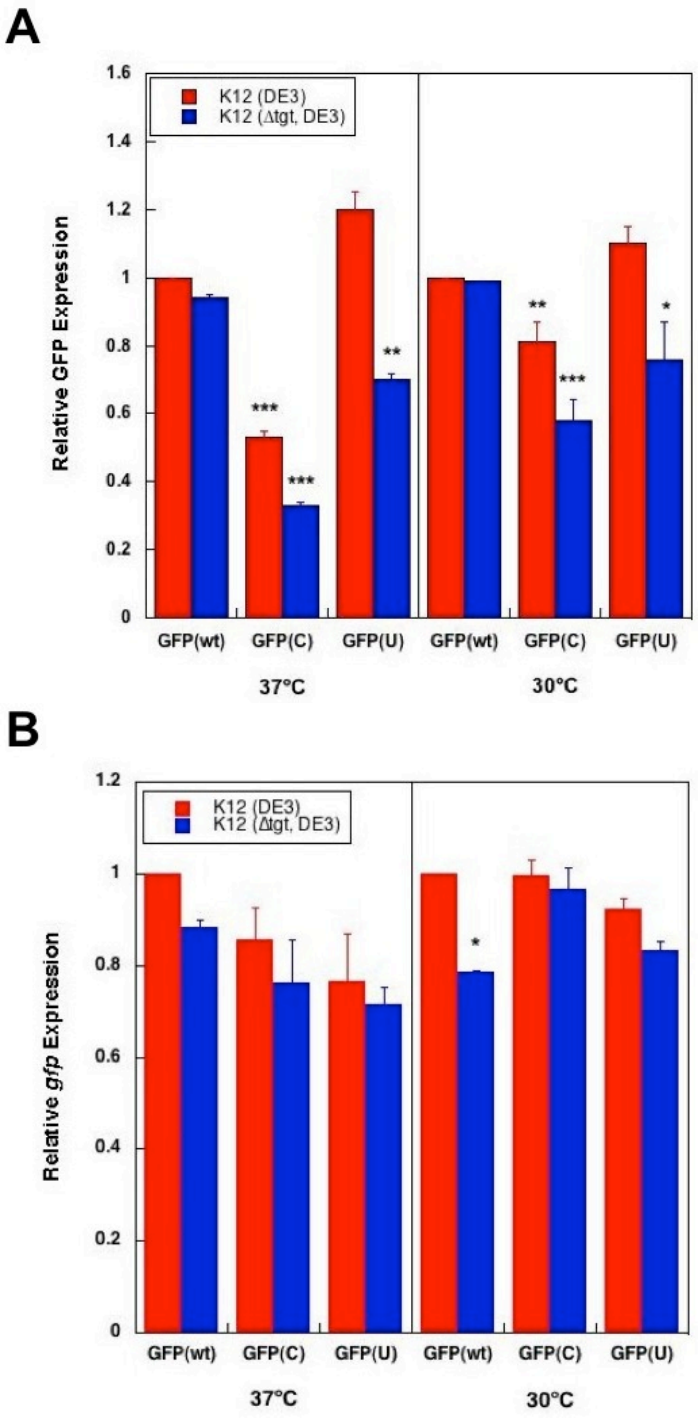


Figure II-4: Relative Expression of Biased-GFP. A) Expression of GFP protein in *E. coli* K12(DE3) versus K12(Δ*tgt*, DE3) at 37ºC and 30ºC relative to GFP(wt) in K12(DE3) at each temperature. B) Expression of *gfp* cDNA by real-time PCR in *E. coli* K12 (DE3) versus K12(Δ*tgt*, DE3) at 37ºC and 30ºC relative to *gfp*(wt) in K12(DE3) at each temperature. Statistical significance determined by Student's t-test: (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.

Table II-3: Comparison of Biased-GFP Expression. The rate of expression for each GFP construct (GFP(wt), C-biased, and U-biased) is listed for both the wildtype K12 (DE3) and K12 (Δ*tgt,* DE3) *E. coli* strains. The average rate of GFP expression is presented with the standard deviation in parentheses.

	Rate of Expression K12 ($F \bullet mg^{-1} \bullet hr^{-1}$)	Rate of Expression K12(Δ*tgt*) ($F \bullet mg^{-1} \bullet hr^{-1}$)	Ratio K12(Δ*tgt*) vs. K12	K12 Relative to GFP(wt) K12	K12(Δ*tgt*) Relative to GFP(wt) K12
Growth @ 37 °C					
GFP(wt)	376 (± 70)	354 (± 70)	0.94	1	0.94 (± 0.01)
GFP(C)	198 (± 30)	124 (± 20)	0.63	0.53 (± 0.02)	0.33 (± 0.008)
GFP(U)	433 (± 100)	289 (± 60)	0.67	1.2 (± 0.05)	0.77 (± 0.02)
Growth @ 30°C					
GFP(wt)	138 (± 20)	137 (± 20)	0.99	1	0.99 (± 0.001)
GFP(C)	112 (± 8)	79.8 (± 20)	0.71	0.81 (± 0.06)	0.58 (± 0.06)
GFP(U)	157 (± 30)	104 (± 30)	0.66	1.1 (± 0.05)	0.76 (± 0.1)

expression relative to GFP(wt) in K12(DE3) at each temperature are represented graphically in Figure II-4A. At both 30°C and 37°C, the rate of GFP expression for either biased open reading frame was approximately 35% slower in the mutant K12(Δ*tgt*, DE3) relative to wild-type K12(DE3). The rate of GFP translation from each codon-biased construct was slower in every case in comparison to wild-type GFP, with the exception of GFP(U) in K12(DE3), however this increased rate of synthesis was not found to be statistically significant (Figure II-4A).

Real-time PCR Analysis of gfp mRNA

E. coli K12(DE3) and K12(Δ*tgt*, DE3) expressing GFP(wt), GFP(C), and GFP(U) were induced with IPTG as above. Cells were harvested following a 3 h

induction period, and total RNA was isolated. Using gene-specific primers, samples were treated with SuperScriptIII™ reverse transcriptase to generate *gfp* cDNA and 16S cDNA as an endogenous control. Generation of *gfp* and 16S cDNA in the same sample was unsuccessful due to the high concentration of 16S rRNA in each total RNA preparation. The concentration of *gfp* cDNA in each sample was determined by averaging the C_T data from each biological replicate, with the exception of *gfp*(wt) in K12(DE3) at 37ºC because one of the duplicate samples was deemed to be an outlier (Figure II-A3-B in Appendix II-2). The concentration of codon-biased *gfp* cDNA was normalized to the concentration of *gfp*(wt) cDNA in K12(DE3) at each temperature (Figure II-4B). The only statistically significant difference in *gfp* mRNA concentration was observed in *gfp*(wt) K12(Δ*tgt*, DE3) at 30ºC. We conclude that the changes we observed in GFP protein expression are predominantly due to changes in the rate of protein synthesis rather than mRNA concentration.

Discussion

The queuine modification is well conserved in eubacterial and eukaryl tRNA, yet the biological function of this modified nucleoside is poorly understood. It is generally believed that the presence of modified bases in tRNA (especially in the anticodon arm) serves to increase the fidelity of the tRNA-mRNA-ribosome interaction during protein synthesis. While studies have been conducted monitoring changes in affinity for Q-cognate tRNAs with the mRNA-ribosome complex in the presence and absence of all modifications, little is known about the contributions made specifically by the queuine nucleoside [9]. It has been

suggested that modification with queuine at the wobble position of the cognate tRNAs leads to preferential codon recognition [10, 11]. However, definitive *in vivo* studies have yet to be conducted to elucidate the effect of queuine, and by extension TGT, in protein translation.

The global expression profiles identified in K12 and K12(Δ*tgt*) by DNA microarray analysis provided insight into some of the functional protein classes affected by the absence of Q-modified tRNA. The identification of trends in genes with decreased (*i.e.,* fimbrial synthesis) and increased (*i.e.,* ribosomal subunits) expression profiles suggests a potential link between queuine and translation of vital cellular components. In rapidly dividing cells, loss of Q-modified tRNA resulted in decreased expression of fimbrial proteins that are critical for cell-cell interaction and generally linked to virulence in pathogenic bacteria. Although no pronounced queuine-cognate codon bias was observed for these open reading frames, the microarray does not necessarily reflect a direct effect on expression profile changes. For example, tRNA-Q_{34} might not directly affect the expression of the fimbrial genes identified. The Δ*tgt* mutation may instead directly impact expression of an upstream factor responsible for expression of these specific genes. Similarly, perhaps an increased expression of ribosomal subunits in stationary phase reflects a cellular response to a deficit in global protein synthesis. It has been demonstrated previously that growth of the *E. coli* (Δ*tgt*) mutant is challenged under hypoxic conditions [12].

To identify genes with pronounced Q-cognate codon bias, the codon usage of genes in the *E. coli* and *S. flexneri* genomes was evaluated with the

biophysical tool CodonW [8]. Genes identified with a pronounced Q-cognate codon bias in *E. coli* and *S. flexneri* revealed a larger subset of genes closely related to those observed by microarray. Interestingly, both the classes of fimbrial proteins and ribosomal subunits emerged in the genome analysis, with the fimbrial proteins biased heavily toward NAU codons and the ribosomal protein subunits containing a bias for NAC codons (Figure II-1). In contrast to this function-specific codon usage, many functional protein classes were shared between the two groups, with several unrelated genes of similar function demonstrating either pronounced NAU- or NAC-bias (*i.e.,* membrane-associated proteins, LPS biosynthesis proteins, and a variety of enzymes). One pattern that was more pronounced in the NAU-bias analysis was the predominance of DNA-binding transcriptional regulators in the category “nucleic acid-binding proteins“. This is interesting because VirF also belongs to this functional class, and it is possible that a slight decrease in expression of these proteins could result in more dramatic effects in expression of their downstream DNA targets.

The bias observed in *S. flexneri* virulence genes for NAU codons is striking. The majority of genes identified belonged to the *mxi-spa* operons, responsible for synthesis of the structural components of the type III secretion system (TTSS) in *Shigella*. The TTSS is a key virulence determinant, and bacterial species deficient in this protein complex are non-invasive [13]. Several secreted proteins (classified as “invasion associated proteins”) were also characterized as strongly NAU-biased. Like the structural TTSS complex, the secreted invasion proteins are also critical for virulence. These results could

suggest a greater extent of regulation of the virulence cascade of *S. flexneri*, where several key components in the invasion process (including *virF*) are strongly NAU-biased and their expression is potentially regulated by the presence of Q-modified tRNA.

To gain insight into the effect of Q-modified tRNA on translation of codon-biased open reading frames, we decided to study expression of a gene with known differential expression in the absence of TGT *in vivo*. We searched for a suitable model protein to study the effects of translation with tRNA-Q_{34} and -G_{34}. Protein would be expressed from candidate mRNA in three forms: 1) with wild-type codon usage, 2) engineered with 100% bias in NAC queuine-cognate codons, and 3) engineered with 100% bias in NAU queuine-cognate codons. Qualifications for a potential candidate included soluble protein, relatively simple quaternary structure, and a simple assay method to determine protein concentration. VirF was a logical choice, as the *virF* mRNA has an overall bias of 80% for NAU queuine-cognate codons, and Durand and colleagues have demonstrated that expression of the VirF protein in the Δ*vacC* mutant (*tgt* homologue in *Shigella*) is reduced to ~40% of the concentration in wild-type *Shigella* [14]. Expression of VirF was very problematic *in vivo* and *in vitro* (see Chapter IV for details), and a suitable alternative gene with pronounced Q-cognate codon bias was difficult to find in the above analyses. While not endogenous to *S. flexneri*, green fluorescent protein (GFP), is widely used in molecular biology, similar in size to VirF and very easy to assay. Although *gfp* mRNA exhibits no pronounced overall bias for Q-cognate codons, the overall

distribution and propensity for clustering of Q-cognate codons in *gfp* mRNA mimics that of the *virF* mRNA (Figure II-3). Therefore GFP was chosen for our study.

GFP expression from wild-type (*gfp*(wt)) and codon-biased constructs (*gfp*(C) and *gfp*(U)) was monitored by fluorescence ($F_{485/518nm}$) and normalized to the weight of the cell pellet in each aliquot to account for differences in cell growth between samples. The total protein content was determined by Bradford assay, and it was found that the amount of protein in each sample did not vary significantly between K12 and K12(Δ*tgt*) (data not shown). This suggests that the differences in GFP expression between K12 and the K12(Δ*tgt*) mutant reflect different rates of GFP synthesis specifically and not overall protein translation. To account for the concern that IPTG-induction at 37°C could result in a dramatically enhanced rate of GFP expression relative to that in *Shigella*, growth at 30°C was performed to slow the overall rate of translation and more closely model a biologically relevant rate of protein synthesis.

GFP expression in K12(DE3) (which contains tRNA-Q_{34}) highlights differential rates of translation for NAC- and NAU-biased mRNA in the presence of Q-modified tRNA (Figure II-4A). For GFP(U), the rate of protein expression is comparable to that observed for GFP(wt) in K12(DE3) at both temperatures. Alternatively, the rate of GFP(C) synthesis is reduced by approximately 45% in K12(DE3) at 37ºC. The relative rates of GFP expression at 30ºC in K12(DE3) follow the same pattern; the rate of expression for GFP(U) ≈ GFP(wt) > GFP(C). In K12(Δ*tgt*, DE3), a different pattern in relative GFP expression is observed. At

both 37ºC and 30ºC, the rate of expression for GFP(wt) > GFP(U) > GFP(C) in the presence of tRNA-G_{34}.

With changes in codon usage of the *gfp* mRNA, we have also changed the G-C content and perhaps the structural dynamics and stability of the codon-biased *gfp* mRNA. To monitor the concentration of *gfp* mRNA following IPTG-induction (3 h), real-time PCR was performed to probe the subsequent effects on RNA stability in each sample. A period of 3 h was chosen because at this stage the variation in total GFP protein levels between samples is the highest, and it provides a direct correlation between mRNA concentration and the rate of translation. The concentration of the endogenous control (16S rRNA) was too high, and subsequent dilution of the reverse transcription samples produced inconsistent data in the real-time PCR analysis (see Appendix II-2 for details). Alternatively, the concentration of codon-biased *gfp* cDNA was normalized to the concentration of *gfp*(wt) cDNA in K12(DE3) at each corresponding temperature (Figure II-4B). This analysis provides for a direct comparison to the levels of GFP protein expressed in each cell line relative to GFP(wt) in *E. coli* K12 as it relates to the total *gfp* mRNA population (Figure II-4A). In general, *gfp* mRNA expression levels are consistent between cell lines and constructs with altered codon usage, as the variability observed was not found to be statistically significant (Figure II-4B). These results suggest that the observed differences in the rate of expression between GFP(wt) and the codon-based GFP proteins is due to changes in the rate of protein translation.

Few experiments have been conducted to directly assess the relevance of queuine-cognate codon bias. *In vivo* analysis with Q-cognate tRNA concluded that tRNA-Q_{34} demonstrated no significant bias between the NAU and NAC codons under competitive conditions with tRNA-G_{34} [10]. A computational model of Q-cognate tRNA codon bias performed with $tRNA^{Asp}$-Q_{34} and $tRNA^{Asp}$-G_{34} predicted the same result, suggesting that the same number of hydrogen-bonds (2) were formed between queuosine and either uridine or cytosine in the wobble position due to an increased hydrogen-bonding network and anticodon rigidity in the presence of queuine [11].

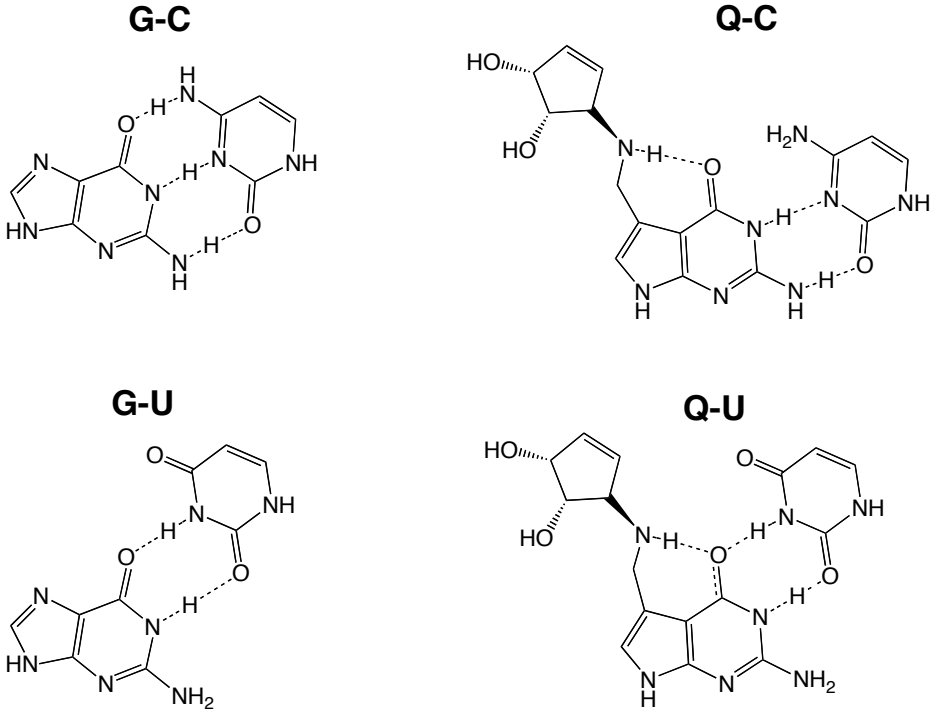


Figure II-5: Predicted Hydrogen Bonding Interactions Formed with Guanine versus Queuine. The four predicted interactions between tRNA-G_{34} and -Q_{34} with cytosine versus uridine in the wobble position are illustrated with the key heterocyclic bases (G/Q with C/U). The intramolecular hydrogen bond between the amino side chain of queuine (Q) and O6 was predicted by molecular dynamics Morris *et al.* (1999), *Journal of Biomolecular Structure and Dynamics*, **16**(4): 757-74.

However, the rate of GFP(U) expression in wild-type *E. coli* K12(DE3) is faster than the rate observed for GFP(C) (100% increase at 37ºC and 30% increase at 30ºC). The observation that the rate of GFP(C) synthesis in K12(Δ*tgt*, DE3) is the slowest for all constructs tested is also in apparent conflict with the originally proposed "codon bias model," which suggested tRNA-G_{34} preferentially recognizes the NAC codon as opposed to the NAU codon [10]. One might predict that if tRNA-G_{34} preferentially recognized the NAC codon, the rate of GFP(C) synthesis would be highest in the (Δ*tgt*) mutant strain.

In some cases, the wobble base modification, or the nucleotide 3' to the wobble position, also exhibits characteristics of codon context (*i.e.,* the interaction between the A-site tRNA and the neighboring codon / P-site tRNA at the ribosome-mRNA complex that can lead to frameshifting) [15]. This phenomenon has been most well-characterized for the wyosine modification at G37 in $tRNA^{Phe}$, which helps prevent frameshifting in recognition of the UUU codon [16, 17]. Using the same computational modeling parameters, Morris *et al*. found no such codon context for the queuine modification in $tRNA^{Asp}$ [11]. Previous studies, however, have suggested an element of codon context with tRNA-Q_{34}, as the tobacco mosaic virus $tRNA^{Tyr}$ (antidocon GΨA) recognizes the UAG stop, but $tRNA^{Tyr}$ (antidocon QΨA) does not [18].

Perhaps the issue of "preferential codon recognition" extends to both tRNA-G_{34} and -Q_{34}. At 30ºC, while the rate of GFP(C) expression in K12(DE3) is relatively indistinguishable from the rate of GFP(U) synthesis in K12(Δ*tgt*), there is a 2-fold difference in the rate of expression of each codon-biased construct in

the alternate cell line (Figure II-4A). This observation suggests that the rate of translation of NAU-biased mRNA is specifically increased with tRNA-Q_{34}, while the rate of translation for NAC-biased mRNA is specifically decreased with tRNA-G_{34} in comparison to mRNA with equivalent usage of the degenerate Q-cognate codons.

It is important to note that when measuring the rate of GFP expression, we are monitoring the overall rate of translation. It is conceivable that preferential binding may relate to the stability of the tRNA-mRNA complex in the A-site or the E-site in the ribosome. In the process of protein synthesis (following the initiation phase with binding of $tRNA^{fMet}$), the appropriate aminoacylated tRNA binds to the corresponding codon in the A-site of the ribosome-mRNA complex [19]. Peptidyl transfer occurs where a new peptide bond is formed between the amino acid attached to the A-site tRNA and the carboxyl end of the polypeptide chain

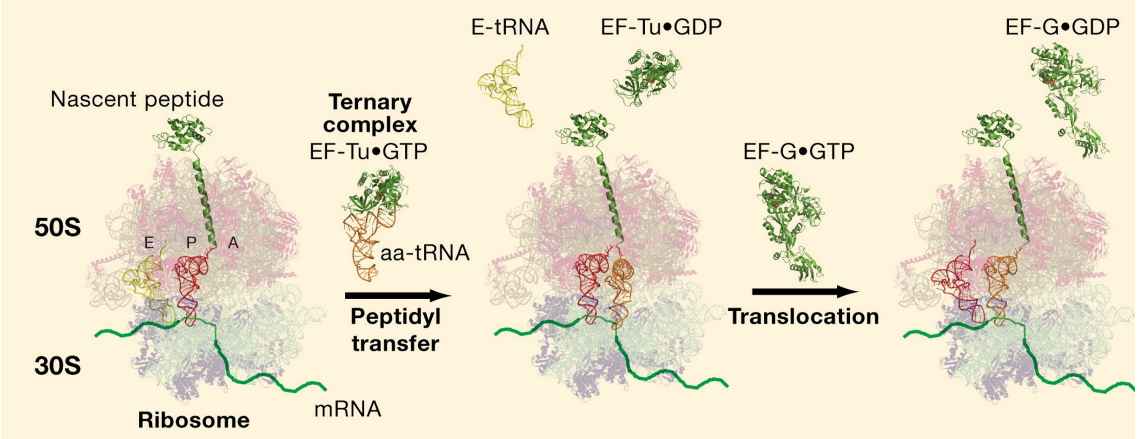


Figure II-6: Snapshot of the Prokaryotic Ribosome in Elongation Phase. During protein translation, tRNA binds to the bacterial ribosome-mRNA complex in three sites: the A-, P-, and E-site. In the elongation phase, the next required aminoacylated tRNA binds to the A-site. Peptidyl transfer occurs, followed by translocation of the ribosome. Adapted from Zaher and Green. (2009), *Cell*, **136**(4): 746-762.

attached to the tRNA in the P-site. The ribosomal subunits shift; the empty tRNA moves from the P-site to the E-site, while the A-site tRNA with the growing polypeptide chain shifts to the P-site (Figure II-6). Once in the E-site, the tRNA dissociates from the ribosome-mRNA complex, and the cycle continues.

If the interaction between tRNA and the cognate codon in the mRNA is too weak at the A-site, the increased on-off rate of the tRNA could halt translation and reduce protein synthesis [9]. Likewise, a strong tRNA-mRNA interaction in the E-site of the ribosome could prevent overall elongation of the polypeptide. Both cases could produce similar results in the total protein analysis, which measures only the end product of protein synthesis. Previous *in vitro* kinetic analysis with tRNA in both fully-modified and fully-unmodified forms suggests that the ultimate role of modified nucleosides in tRNA is to normalize the interaction at the ribosome-mRNA complex [9]. It is conceivable that codon usage in a given mRNA performs the same function of normalization between the anticodon-codon interactions, and any changes in the already optimized process lead to unfavorable results. Alternatively, the changes in queuine-cognate codon usage may alter the rates of translation in ways we do not fully understand under the current experimental protocol.

Taken together, we propose a new interpretation of the queuine-cognate "codon bias" model. Perhaps the strength of the anticodon-codon interaction (*i.e.,* 3 full Watson-Crick hydrogen bond pairs or 2 hydrogen bonds) partially determines the rate of protein synthesis. The slowest rate of protein expression at both 30°C and 37°C was observed in the only case where a full Watson-Crick

interaction at the wobble position is formed: GFP(C) expressed in K12(Δ*tgt*, DE3). The similarity between rate of expression for GFP(U) with tRNA-G_{34} and GFP(C) with tRNA-Q_{34} may also suggest a component of codon preference. Two hydrogen bonds are formed in both cases, yet studies with Q-modified tRNA show an increased rate of translation for GFP(U) and GFP(wt) utilizing the same number of hydrogen bonds in the anticodon-codon interaction. This result could be predicted if tRNA-Q_{34} formed an optimized interaction with NAU codons that was different from the interaction formed by tRNA-G_{34}. Although the effect in GFP expression is subtle, the result could be more dramatic for proteins that must maintain a specific cellular concentration for proper activity, as is the case with VirF.

Conclusions

The modification state of tRNA with queuine appears to have pronounced effects on global gene expression profiles in *E. coli*. Although tRNA-G_{34} and -Q_{34} can translate both degenerate Q-cognate codons (NAU and NAC), the rate of protein synthesis is altered with each tRNA-codon pair. The presence of queuine/guanine in cognate tRNA may provide a different anticodon-codon interaction with each degenerate codon. This effect could help explain the observation that *S. flexneri* VirF protein levels (translated from NAU-biased *virF* mRNA) are reduced in the (Δ*vacC*) mutant. Further studies must be performed to explore the importance of Q-tRNA in the translation of other genes critical to virulence and cellular structure/growth.

Notes to Chapter II

We gratefully acknowledge Dr. Taocong Jin in the Molecular Biology Core Facility (University of Michigan Dental School) for his assistance in experiments and data analysis related to DNA microarray and real-time PCR. We thank Dr. Laura Miesle (Sardelli) for preparation of the total *E. coli* RNA for DNA microarray analysis and Dr. Adam Lee for his preliminary analysis of the CodonW data. We also thank Prof. Katrin Karbstein for use of her lab supplies in the Northern blotting analysis, and Crystal Young for her instruction on the technique.

Abbreviations used: TGT, tRNA-guanine transglycosylase; Q, queuosine; IPTG, isopropyl β-D-thiogalactoside; dNTP, deoxyribonucleotide triphosphate; GFP, green fluorescent protein; DTT, dithiothreitol; RT, reverse transcription; HPLC, high performance liquid chromatography; C_T, threshold cycle for target amplification in real-time PCR; PCR, polymerase chain reaction; TTSS, type III secretion system.

Appendix II-1

Northern Blot Analysis of *gfp* mRNA

Analysis of RNA has historically been performed by Northern blotting with single-stranded oligonucleotides that are antisense to the target RNA. Using a modified protocol developed in the Karbstein lab (University of Michigan), I attempted to analyze the concentration of *gfp* mRNA by Northern blot using both an active and passive transfer method. The active transfer of *gfp* mRNA (~700 nt) from a denaturing acrylamide gel was unsuccessful, as the technique is best suited for RNA ≤ 500 nt in length. For passive transfer, total RNA (10 µg) was loaded onto a 1.25% agarose gel (with 6.7% formaldehyde) in MESA buffer (20 mM MOPS (pH 7), 8 mM sodium acetate, 1 mM EDTA). Samples were run at 95 V for 7 h. Following electrophoresis, the gel was treated with 50 mM NaOH for 20 min. at room temperature with shaking. The gel was incubated in ethidium bromide solution (0.0001% EtBr in 0.5X TBE buffer (5 g Tris Base (pH 8.0), 2.8 g boric acid, 1 mM EDTA / 1 L dH_2O) for 30 min. at room temperature followed by a destain solution (0.5X TBE) for an additional 30 min. to visualize the total RNA (Figure II-A1).

The total RNA was transferred to the Hybond-N membrane in SSC buffer (3 M NaCl, 300 mM sodium citrate (pH 7)) overnight according to the vendor's protocol (Amersham). Oligonucleotide probes for *gfp* mRNA (5'–CACTTGTACAGCTCGTCCATGCCAAG–3') and 23S rRNA as an endogenous control (5'–AAGGTTAAGCCTCACGGTTC–3') were labeled with ^{32}P-γ-ATP using

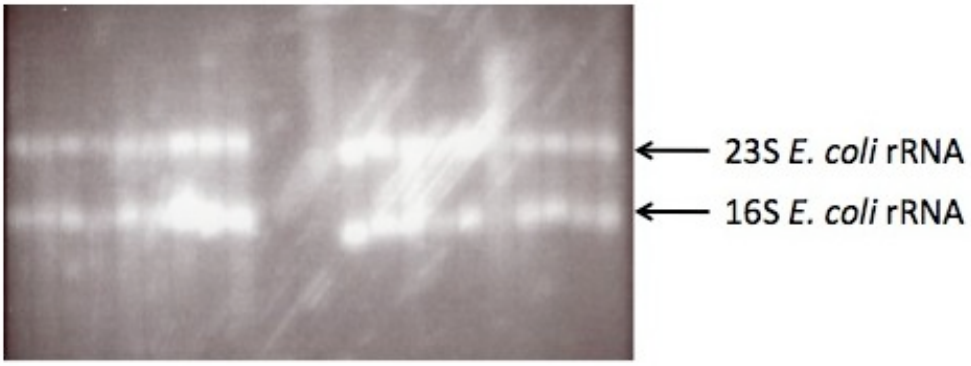


Figure II-A1: Denaturing Agarose Gel for *gfp* Northern Blot. Ethidium bromide staining of total RNA with induced *gfp* mRNA. All samples were run in duplicate: *gfp*(wt), *gfp*(C), and *gfp*(U) in K12(DE3) and K12(Δ*tgt*, DE3), induced at 37ºC and 30ºC.

T4 polynucleotide kinase (PNK) according to the vendor's protocol (New England Biolabs). Free NTPs were removed with G25 microspin columns (GE Healthcare), and the specific activity was determined (1 µL probe) by scintillation counting. The membrane was cross-linked, and treated with boiling 0.1% SDS (sodium dodecyl sulfate) at 65ºC with rotation for 10 min. Pre-hybridization solution was added (16 mL total volume; Denhardt's solution (0.15% Ficoll (type 400), 0.15% polyvinylpyrrolidone, 0.15% BSA); SSPE buffer (750 mM NaCl, 50 mM NaH_2PO_4 (pH 7.4), 6.25 mM EDTA); 0.1% SDS; boiled salmon sperm DNA (790 µL)) at 37ºC for 4 h with rotation. The *gfp* probe (1 x 10^7 counts) was boiled with 500 µL pre-hybridization solution for 3 min., and the labeled probe was incubated with the membrane at 37ºC overnight.

The membrane was washed with subsequent volumes of Solution 1 (25 mL volume, SSPE buffer with 0.1% SDS) to strip unbound probe until the wash volumes contained no detectable levels of radioactivity. The membrane was

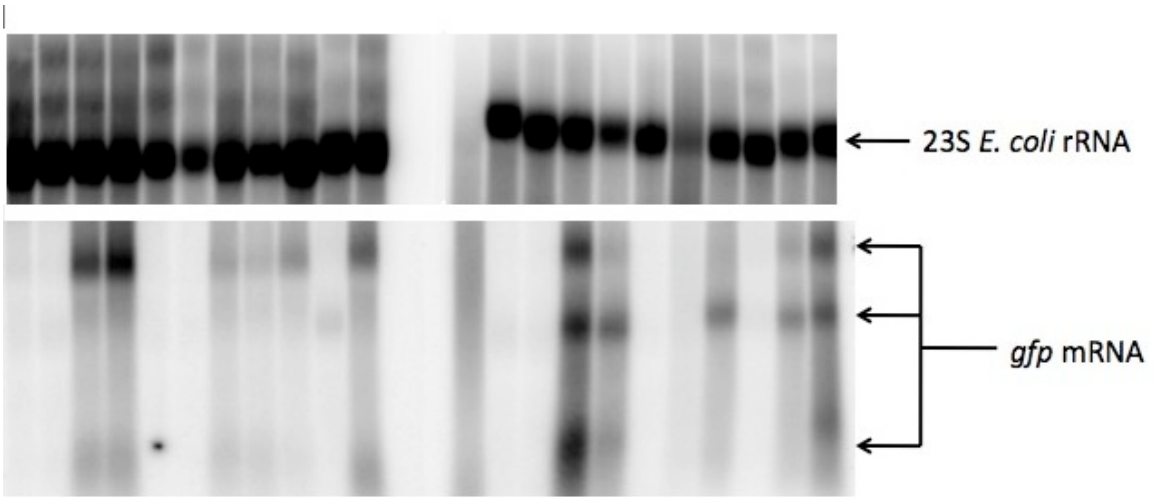


Figure II-A2: Northern Blot Analysis. Total RNA samples probed with ^{32}P-γ-ATP labeled oligonucleotides for *gfp* mRNA and 23S rRNA. 23S rRNA control is fairly consistent, but inconsistent results obtained with *gfp* mRNA. Possible lower molecular weight *gfp* degradation products also marked.

dried and exposed to an x-ray screen for 8 h at room temperature. The screen was analyzed by the Typhoon™ imager using ImageQuant™ software (GE Healthcare). The *gfp* probe was stripped, and the 23S probe was annealed as described previously. Representative membrane blots are shown in Figure II-A2. Detection of 23S rRNA was more successful due to the relative abundance of the RNA, but detection of *gfp* mRNA was much more variable and unreliable.

Appendix II-2

Real-time PCR Analysis Raw Data

Samples for analysis of *gfp* mRNA by real-time PCR were prepared in duplicate. Experiments for detection of nucleic acids by CybrGreen dye were performed as described in Materials and Methods. The raw expression data for the GFP protein in addition to the raw C_T (threshold cycle) data for *gfp* and 16S RNA in each set of biological replicates is illustrated in Figure II-A3. The concentration of 16S rRNA in four samples at 37ºC appears to be very dilute in comparison to the endogenous control in other samples (*gfp*(wt) K12-2, *gfp*(wt) K12(Δ*tgt*)-1, *gfp*(C) K12(Δ*tgt*)-1, *gfp*(U) K12-1) (Figure II-A3-C). These samples were treated as outliers, but the relative consistency in 16S rRNA concentration among the other samples suggests that any variability in *gfp* mRNA conentration is not the result of inconsistent reverse transcription between samples. The variability in threshold cycle for detection of 16S is most likely due to the alterations in protocol (*i.e.,* dilution of RT sample, increased primer concentration) between the endogenous control and *gfp* analysis. The reverse transcription reactions were optimized to the concentration of *gfp* mRNA, and does not account for the abundance of the 16S rRNA. The inconsistency in the 16S control masks the reproducibility in the *gfp* mRNA so the data measured from duplicate samples was instead averaged and reported as expression relative to *gfp*(wt) mRNA in wild-type K12(DE3) at each temperature (Figure II-4B in Results). In general, the reproducibility between replicates was high.

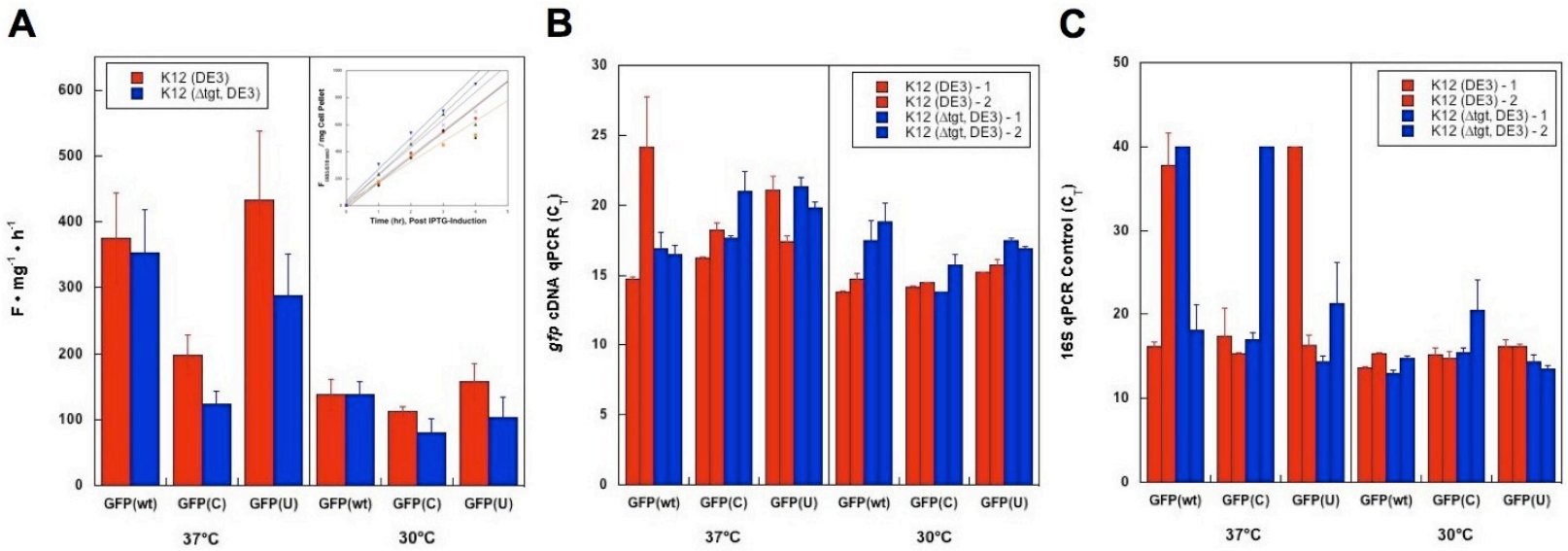


Figure II-A3: Relative Expression of Biased-GFP. A) Expression of GFP protein in *E. coli* K12(DE3) versus K12(Δ*tgt*, DE3) at 37ºC and 30ºC relative to GFP(wt) in K12(DE3) at each temperature. B) Expression of *gfp* cDNA by real-time PCR in *E. coli* K12 (DE3) versus K12(Δ*tgt*, DE3) at 37ºC and 30ºC relative to *gfp*(wt) in K12(DE3) at each temperature.

The one major discrepancy was the level of *gfp*(wt) in K12(DE3) at 37ºC, which had a representative standard deviation of approximately 10 C_T between the duplicate samples (Figure II-A3-B). Because the average of these samples would skew the normalization, and the *gfp*(wt)-2 concentration is likely an outlier, only the concentration of *gfp*(wt)-1 in K12(DE3) was used to normalize the codon biased *gfp* samples at 37ºC. Overall, this method of normalization is a better fit considering the conditions of the experiment.

Appendix II-3

Presentation of Genes Identified by Genome Analysis

E. coli Genome (4149 protein-coding genes)

NAU Codon Bias – 148 open reading frames

Gene Locus Tag	Gene ID	Function/Description
Membrane-associated Proteins		
b3921	yiiR	Conserved inner membrane protein
b3511	hdeD	Acid-resistance membrane protein
b1322	ycjF	Conserved inner membrane protein
b1607	ydgC	Conserved inner membrane protein
b1626	ydgK	Conserved inner membrane protein
b3563	yiaB	Conserved inner membrane protein
b3817	yigF	Conserved inner membrane protein
b0111	ampE	Predicted inner membrane protein
b1447	ydcZ	Predicted inner membrane protein
b2145	yeiS	Predicted inner membrane protein
b2668	ygaP	Predicted inner membrane protein
b3818	yigG	Predicted inner membrane protein
b2666	yqaE	Predicted membrane protein
b1615	uidC	Predicted outer membrane porin
b1038	csgF	Predicted transport protein
LPS Biosynthesis Proteins		
b2563	acpS	Holo-ACP Synthase I
b3623	waaU	Lipopolysaccharide core biosynth.
b3624	rfaZ	Lipopolysaccharide core biosynth.
b3629	rfaS	Lipopolysaccharide core biosynth.
b3622	rfaL	O-antigen ligase
b2035	rfc	O-antigen polymerase
b2371	yfdE	Predicted CoA-transferase, NAD(P)-
Nucleic Acid-binding Protein		
b1320	ycjW	Predicted DNA-binding
b1343	dbpA	ATP-dependent 23s rRNA helicase
b0546	ybcM	DLP12 prophage, predicted DNA-
b1399	paaX	DNA-binding transcriptional
b3512	gadE	DNA-binding transcriptional
b4046	zur	DNA-binding transcriptional
b3149	diaA	DnaA initiator-associated factor for
b2038	rfbC	dTDP-4-deoxyrhamnose-3,5-

b3998	nfi	Endonuclease V
b3856	mobB	Molybdopterin-guanine dinucleotide
b0483	ybaQ	Predicted DNA-binding
b1434	ydcN	Predicted DNA-binding
b2398	yfeC	Predicted DNA-binding
b3060	ttdR	Predicted DNA-binding
b3507	dctR	Predicted DNA-binding
b3180	yhbY	Predicted DNA-binding protein
b0318	yahD	Predicted transcriptional regulator
b3889	yiiE	Predicted transcriptional regulator
b2573	rpoE	RNA polymerase, sigma factor E
b3627	rfaI	UDP-D-galactose:(glycosyl)

Fimbrial / Pilin proteins

b1940	fliH	Flagellar biosynthesis protein
b1942	fliJ	Flagellar protein
b4320	fimH	Minor component of type I fimbriae
b0531	sfmC	Pilin chaperone, periplasmic
b1043	csgC	Predicted curli production protein
b0938	ycbQ	Predicted fimbrial-like adhesion
b2321	flk	Predicted flagella assembly protein
b1546	tfaQ	Qin prophage, predicted tail fibre
b1373	tfaR	Rac prophage, predicted tail fibre

Predicted/Conserved Proteins

b0956	ycbG	Conserved protein
b1705	ydiE	Conserved protein
b1707	ydiV	Conserved protein
b1787	yeaK	Conserved protein
b1878	flhE	Conserved protein
b1976	mtfA	Conserved protein
b2012	yeeD	Conserved protein
b2073	yegL	Conserved protein
b2583	yfiP	Conserved protein
b3098	yqjD	Conserved protein
b3100	yqjK	Conserved protein
b3348	slyX	Conserved protein
b3596	yibG	Conserved protein
b3698	yidB	Conserved protein
b3717	cbrC	Conserved protein
b4594	ymgJ	Hypothetical protein
b4604	yojO	Hypothetical protein
b4621	yjbS	Hypothetical protein
b2301	yfcF	Predicted enzyme
b0329	yahO	Predicted protein
b0466	ybaM	Predicted protein
b0498	ybbC	Predicted protein
b1001	yccE	Predicted protein
b1058	yceO	Predicted protein

b1112	bhsA	Predicted protein
b1167	ymgC	Predicted protein
b1455	yncH	Predicted protein
b1532	marB	Predicted protein
b1847	yebF	Predicted protein
b2085	yegR	Predicted protein
b2274	yfbO	Predicted protein
b2602	yflL	Predicted protein
b2757	ygcI	Predicted protein
b2758	ygcJ	Predicted protein
b2849	yqeK	Predicted protein
b2940	yqgC	Predicted protein
b3107	yhaL	Predicted protein
b3120	yhaB	Predicted protein
b3362	yhfG	Predicted protein
b4026	yjbE	Predicted protein
b4047	yjbL	Predicted protein
b4482	yigE	Predicted protein
b4555	yicS	Predicted protein
Amino Acid-binding Proteins		
b1325	ycjG	L-Ala-D/L-Glu epimerase
b2743	pcm	L-isoaspartate protein
b1838	pphA	Serine/threonine-specific protein
b0367	tauC	Taurine transporter subunit
b2851	ygeG	Predicted chaperone
Enzymes		
b2462	eutS	Predicted carboxysome (carbon
b1270	btuR	Cob(I)alamin adenosyltransferase/
b1992	cobS	Cobalamin 5'-phosphate synthase
b1557	cspB	Qin-prophage, cold-shock protein
b1646	sodC	Superoxide dismutase, Cu, Zn
b2311	ubiX	3-octaprenyl-4-hydroxybenzoate
b1538	dcp	Dipeptidyl carboxypeptidase II
b4367	fhuF	Ferric iron reductase involved in
b2091	gatD	Galactitol-1-phosphate
b2093	gatB	Galactitol-specific enzyme IIB
b3631	rfaG	Glucosyltransferase I
b0849	grxA	Glutaredoxin 1, redox coenzyme w/
b1379	hslJ	Heat-inducible protein
b2259	pmrD	Polymyxin resistance protein B
b2844	yqeF	Predicted acyltransferase
b3765	yifB	Predicted bifunctional enzyme and
b2371	yfdE	Predicted CoA-transferase, NAD(P)-
b2301	yfcF	Predicted enzyme
b3953	frwD	Predicted enzyme IIB component of
b1710	btuE	Predicted glutathione peroxidase

NAC Codon Bias – 95 open reading frames

Gene Locus Tag	Gene ID	Function/Description
Ribosomal Proteins		
b3165	rpsO	30S ribosomal subunit protein S15
b0169	rpsB	30S ribosomal subunit protein S2
b3314	rpsC	30S ribosomal subunit protein S3
b3984	rplA	50S ribosomal subunit protein L1
b2185	rplY	50S ribosomal subunit protein L25
b3320	rplC	50S ribosomal subunit protein L3
b3339	tufA	Protein chain elongation factor EF-Tu
b3980	tufB	Protein chain elongation factor EF-Tu
b0911	rpsA	30S ribosomal subunit protein S1
b3342	rpsL	30S ribosomal subunit protein S12
b2609	rpsP	30S ribosomal subunit protein S16
b3311	rpsQ	30S ribosomal subunit protein S17
b0023	rpsT	30S ribosomal subunit protein S20
b3065	rpsU	30S ribosomal subunit protein S21
b3230	rpsI	30S ribosomal subunit protein S9
b3310	rplN	50S ribosomal subunit protein L14
b3301	rplO	50S ribosomal subunit protein L15
b2606	rplS	50S ribosomal subunit protein L19
b1716	rplT	50S ribosomal subunit protein L20
b3318	rplW	50S ribosomal subunit protein L23
b3637	rpmB	50S ribosomal subunit protein L28
b1089	rpmF	50S ribosomal subunit protein L32
b1717	rpmI	50S ribosomal subunit protein L35
b4147	efp	Elongation factor EF-P
b3340	fusA	Protein chain elongation factor EF-G,
b0170	tsf	Protein chain elongation factor EF-Ts
b0884	infA	Translation initiation factor IF-1
LPS Biosynthesis Proteins		
b3255	accB	Acetyl CoA carboxylase, BCCP subunit
b2207	napD	Assembly protein for periplasmic nitrat
b4093	phnO	Pred. acyltransferase with acyl-CoA
b0729	sucD	Succinyl-CoA synthetase, NAD(P)-bind
b4093	phnO	Predicted acyltransferase with acyl-CoA
Nucleic Acid-binding Proteins		
b2699	recA	DNA strand exchange and recomb.
b3560	glyQ	Glycine tRNA synthetase, alpha subunit
b3871	typA	GTP-binding protein
b4384	deoD	Purine-nucleoside phosphorylase
b3988	rpoC	RNA polymerase, beta prime subunit
b3783	rho	Transcription termination factor
b2498	upp	Uracil phosphoribosyltranserase

Membrane-associated Proteins		
b0814	ompX	Outer membrane protein
b3736	atpF	F0 sector of membrane-bound ATP syn
b3732	atpD	F1 sector of membrane-bound ATP syn
b3731	atpC	F1 sector of membrane-bound ATP syn
b2215	ompC	Outer membrane porin protein C
b0957	ompA	Outer membrane protein A
b1632	rsxE	Predicted inner membrane NADH-quino
b4354	yjiY	Predicted inner membrane protein
Amino Acid-binding Proteins		
b2796	sdaC	Predicted serine transporter
b4143	groL	Cpn60 chaperonin GroEL, large subunit
b3349	slyD	FKBP-type peptidyl prolyl cis-trans iso.
b0810	ginP	Glutamine transporter subunit
b3556	cspA	Major cold shock protein
b0441	ppiD	Peptidyl-prolyl cis/trans isomerase rot.
b0525	ppiB	Peptidyl-prolyl cis/trans isomerase rot.
b0436	tig	Peptidyl-prolyl cis/trans isomerase
b0161	degP	Serine endoprotease (protease Do),
b2551	glyA	Serine hydroxymethyltransferase
Enzymes		
b0605	ahpC	Alkyl hydroperoxide reductase, C22 sub
b0270	yagG	CP4-6 prophage; predicted sugar trans.
b0271	yagH	CP4-6 prophage; predicted xylosidase/
b2935	tktA	Transketolase 1, thiamin-binding
b3919	tpiA	Triosephosphate isomerase
b4196	ulaD	3-keto-L-gulonate 6-phosphate decarb.
b2296	ackA	Acetate kinase A and propionate kinase
b4101	phnG	Carbon-phosphorus lyase complex sub.
b2779	eno	Enolase
b2528	iscA	FeS cluster assembly protein
b2925	fbaA	Fructose-bisphosphate aldolase, class II
b1779	gapA	Glyceraldehyde-3-phosphate dehydrog.
b2023	hisH	Imidazole glycerol phosphate synthase,
b4226	ppa	Inorganic pyrophosphatase
b0116	lpd	Lipoamide dehydrogenase, E3 compon.
b1818	manY	Mannose-specific enzyme IIC compon.
b1850	eda	Multifunctional 2-keto-3-deoxygluconat
b2297	pta	Phosphate acetyltransferase
b2926	pgk	Phosphoglycerate kinase
b1207	prs	Phosphoribosylpyrophosphate synthase
b4565	sgcB	Predicted enzyme IIB component PTS
b0903	pflB	Pyruvate formate lyase I
b0008	talB	Transaldolase B

Predicted/Conserved Proteins		
b0471	ybaB	Conserved protein
b3096	yqjB	Conserved protein
b1541	ydfZ	Conserved protein

S. flexneri Genome (4177 protein-coding genes)

NAU Codon Bias – 111 open reading frames

Gene Locus Tag	Gene ID	Function/Description
Membrane-associated Proteins		
SF2171	yehB	Putative outer membrane protein
SF2676	smpA	Small membrane protein A
SF3545	hdeD	Acid-resistance membrane protein
LPS Biosynthesis Proteins		
SF2100	rfbE	Polysaccharide biosynthesis protein
SF2625	acpS	4'-phosphopantetheinyl transferase
SF3665	waaY	LPS core biosynthesis protein
Nucleic Acid-binding Proteins		
SF1826	dbpA	ATP-dependent RNA helicase DbpA
SF3190	yraO	DnaA initiator-associating protein
SF3927	mobB	Molybdopterin-guanine dinucleotide
SF1326	ycjW	Putative LACI-type transcriptional
SF2906	ygfI	Putative LYSR-type transcriptional
SF2635	rpoE	RNA polymerase sigma factor RpoE
SF3220	yhbY	RNA-binding protein
SF2016	yedW	Transcriptional regulatory protein
Fimbrial/Pilin Proteins		
SF1891	--	Putative tail component of prophage CP-93
SF1922	flhE	Flagellar protein
SF1987	fliJ	Flagellar protein
SF1895	--	Putative tail attachment protein
SF1890	--	Putative tail component of prophage
Predicted Proteins		
SF1005	yccE	Hypothetical Protein SF1005
SF1437	yeaK	Hypothetical Protein SF1437
SF3538	--	Hypothetical Protein SF3538
SF4232	yjgN	Hypothetical Protein SF4232
SF0890	ybhH	Hypothetical protein SF0890
SF0913	ycaR	Hypothetical protein SF0913
SF0916	ycbC	Hypothetical protein SF0916
SF1065	yceO	Hypothetical protein SF1065
SF1116	ycfR	Hypothetical protein SF1116
SF1169	--	Hypothetical protein SF1169
SF1278	--	Hypothetical protein SF1278
SF1328	ycjF	Hypothetical protein SF1328
SF1333	--	Hypothetical protein SF1333

SF1336	--	Hypothetical protein SF1336
SF1498	ydjO	Hypothetical protein SF1498
SF1524	--	Hypothetical protein SF1524
SF1562	marB	Hypothetical protein SF1562
SF1628	ydgC	Hypothetical protein SF1628
SF1752	yddG	Hypothetical protein SF1752
SF1770	--	Hypothetical protein SF1770
SF1858	yebF	Hypothetical protein SF1858
SF2071	yeeD	Hypothetical protein SF2071
SF2093	--	Hypothetical protein SF2093
SF2136	--	Hypothetical protein SF2136
SF2137	--	Hypothetical protein SF2137
SF2230	--	Hypothetical protein SF2230
SF2377	yfcF	Hypothetical protein SF2377
SF2438	yfdE	Hypothetical protein SF2438
SF2460	yfeC	Hypothetical protein SF2460
SF2486	--	Hypothetical protein SF2486
SF2645	yfiP	Hypothetical protein SF2645
SF2662	yfiL	Hypothetical protein SF2662
SF2932	yqgB	Hypothetical protein SF2932
SF2987	--	Hypothetical protein SF2987
SF3007	--	Hypothetical protein SF3007
SF3141	yqjD	Hypothetical protein SF3141
SF3143	--	Hypothetical protein SF3143
SF3149	yhaL	Hypothetical protein SF3149
SF3279	yhcO	Hypothetical protein SF3279
SF3366	slyX	Hypothetical protein SF3366
SF3381	yhfG	Hypothetical protein SF3381
SF3421	--	Hypothetical protein SF3421
SF3500	--	Hypothetical protein SF3500
SF3503	yhiI	Hypothetical protein SF3503
SF3541	yhiF	Hypothetical protein SF3541
SF3546	yhiE	Hypothetical protein SF3546
SF3607	yiaB	Hypothetical protein SF3607
SF3766	yidB	Hypothetical protein SF3766
SF3895	yigF	Hypothetical protein SF3895
SF3902	yigK	Hypothetical protein SF3902
SF3963	--	Hypothetical protein SF3963
SF3965	yiiE	Hypothetical protein SF3965
SF4158	yjbL	Hypothetical protein SF4158
SF4179	yjbE	Hypothetical protein SF4179
SF4195	yjiC	Hypothetical protein SF4195
SF4417	--	Hypothetical protein SF4417

Enzymes

SF2387	ubiX	3-octaprenyl-4-hydroxybenzoate carboxylase
SF1272	btuR	Cob(I)yrinic acid a,c-diamide adenosyltransf.
SF2060	cobS	Cobalamin synthase
SF2099	rfbF	dTDP-rhamnosyl transferase

SF2096	rfbI	Glycosyl translocase
SF2625	acpS	4'-phosphopantetheinyl transferase
SF1578	--	Aldolase
SF0990	cspH	Cold shock-like protein
SF2101	rfbC	dTDP-6-deoxy-L-mannose-
SF2098	rfbG	dTDP-rhamnosyl transferase
SF4070	nfi	Endonuclease V
SF4400	fhuF	Ferric iron reductase involved in ferr
SF2150	gatR	Galactitol utilization operon
SF2151	gatD	Galactitol-1-phosphate dehydrogena
SF2155	gatB	Galactitol-specific PTS system
SF0802	grxA	Glutaredoxin 1
SF2096	rfbJ	Glycosyl transferase
SF1789	trg	Methyl-accepting chemotaxis protei
SF2766	pcm	Protein-L-isoaspartate O-MTase
SF4377	hpaA	Putative 4-hydroxyphenylacetate
SF2781	ygcM	Putative 6-pyruvoyl tetrahydro-
SF1521	btuE	Putative glutathione peroxidase
SF1331	ycjG	Putative muconate cycloisomerase I
SF1673	sodC	Superoxide dismutase

NAC Codon Bias – 84 open reading frames

Gene Locus Tag	Gene ID	Function/Description
Ribosomal Proteins		
SF3206	rpsO	30S ribosomal protein S15
SF3346	rpsC	30S ribosomal protein S3
SF4057	rplA	50S ribosomal protein L1
SF2272	rplY	50S ribosomal protein L25
SF3352	rplC	50S ribosomal protein L3
SF3357	tuf	Elongation factor Tu
SF4053	tuf	Elongation factor Tu
SF0907	rpsA	30S ribosomal protein S1
SF3360	rpsL	30S ribosomal protein S12
SF2669	rpsP	30S ribosomal protein S16
SF3343	rpsQ	30S ribosomal protein S17
SF3106	rpsU	30S ribosomal protein S21
SF3270	rpsI	30S ribosomal protein S9
SF4056	rplK	50S ribosomal protein L11
SF3342	rplN	50S ribosomal protein L14
SF2666	rplS	50S ribosomal protein L19
SF1515	rplT	50S ribosomal protein L20
SF3350	rplW	50S ribosomal protein L23
SF3676	rpmB	50S ribosomal protein L28
SF1093	rpmF	50S ribosomal protein L32
SF1514	rpmI	50S ribosomal protein L35
SF3333	rplO	50S ribosomal protien L15
SF4303	efp	Elongation factor P
SF3358	fusA	GTP-binding protein chain elong. Factor EF-
SF0843	infA	Translation initiation factor IF-1
LPS Biosynthesis Proteins		
SF3293	accB	Acetyl-CoA carboxylase biotin carrier protein
SF2291	napD	Assembly protein for periplasmic nitrate red.
Nucelic Acid-binding Proteins		
SF4061	rpoC	DNA-directed RNA polymerase subunit β'
SF3603	glyQ	Glycyl-tRNA synthetase subunit α
SF4416	deoD	Purine nucleoside phosphorylase
SF2542	upp	Uracil phosphoribosyltransferase
Membrane-associated Proteins		
SF1706	lpp	Murein lipoprotein
SF0957	ompA	Outer membrane protein A
SF0765	ompX	Outer membrane protein X
SF0761	glnP	Glutamine ABC transporter permease protein
SF2299	ompC	Outer membrane porin protein C

Hypothetical Proteins		
SF1599	--	Hypothetical protein
SF2274	yejL	Hypothetical protein
SF3139	yqjB	Hypothetical protein
SF1554	--	Hypothetical protein
SF2967	--	Hypothetical protein
Amino Acid-binding Proteins		
SF2810	sdaC	Serine Transporter
SF4297	groEL	Chaperonin GroEL
SF3850	ppiC	Peptidyl-prolyl cis-trans isomerase C
Enzymes		
SF3997	tpiA	Triosephosphate isomerase
SF2372	ackA	Acetate kinase
SF3816	atpF	F0F1 ATP synthase subunit B
SF3812	atpD	F0F1 ATP synthase subunit beta
SF3811	atpC	F0F1 ATP synthase subunit epsilon
SF0898	pflB	Formate acetyltransferase 1
Sf2910	fba	Fructose-bisphosphate aldolase
SF1444	gapA	Glyceraldehyde-3-phosphate dehydrogenase
SF1860	eda	keto-hydroxyglutarate-aldolase/
SF2373	pta	Phosphate acetyltransferase
SF3729	pstA	Phosphate transporter permease subunit
SF2911	pgk	Phosphoglycerate kinase
SF1410	manY	PTS enzyme IIC
SF1904	--	Putative serine protease
SF4024	--	Putative transaldolase
SF2722	recA	Recombinase A
SF1210	prsA	Ribose-phosphate pyrophosphokinase
SF1657	ydgQ	SoxR-reducing system protein RsxE
SF2358	nuoH	NADH dehydrogenase subunit H
SF2356	nuoJ	NADH dehydrogenase subunit J

S. flexneri Virulence Plasmid (293 protein-coding genes)

NAU Codon Bias – 24 open reading frames

Gene Locus Tag	Gene ID	Function/Description
TTSS Genes		
pWR501_p155	mxiA	Invasion protein
pWR501_p161	spaP	Surface presentation of antigens protein SpaP
pWR501_p147	mxiJ	Type III secretion protien
pWR501_p149	mxiL	Hypotetical protein
pWR501_p150	mxiM	Hypotetical protein
pWR501_p166	spa-orf11	Hypotetical protein
pWR501_p154	mxiC	Invasion protein
pWR501_p158	spa13	Invasion protein
pWR501_p151	mxiE	Putative lipoprotein
pWR501_p148	mxiK	Type III secretion protein
pWR501_p152	mxiD	Type III secretion protein
pWR501_p153	mxiD	Type III secretion protein
Invasion Proteins		
pWR501_p192	--	Adhesion protein, fragment
pWR501_p142	ipgE	Invasion-associated protein
pWR501_p143	ipgF	Invasion-associated protein
pWR501_p237	mob9	Plasmid mobilization protein
pWR501_p028	--	Putative transposase, fragment
pWR501_p141	ipgD	Secreted protein
Nucleic Acid-binding Protein		
pWR501_p051	virF	Transcriptional activator of virulence loci
Hypothetical Proteins		
pWR501_p120	--	Hypotetical protein
pWR501_p236	--	Hypotetical protein
pWR501_p264	yigB	Hypotetical protein
pWR501_p290	--	Hypotetical protein
pWR501_p266	--	IS600 orfA, fragment

1. Welch, M., et al., *You're one in a googol: optimizing genes for protein expression.* Journal of the Royal Society Interface, 2009. **6**: p. 10.

2. Agris, P.F., *Decoding the genome: a *modified* view.* Nucleic Acids Research, 2004. **32**(1): p. 223-238.

3. Grosjean, J.H., S.D. Henau, and D.M. Crothers, *On the physical basis for ambiguity in genetic coding interactions.* Proceedings of the National Academy of Sciences, 1978. **75**(2): p. 610-614.

4. Kudla, G., et al., *Coding-Sequence Determinants of Gene Expression in Escherichia coli.* Science, 2009. **324**(5924): p. 255-258.

5. Durand, J.M., et al., *Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of Shigella flexneri: the expression of the virF gene.* Molecular Microbiology, 2000. **35**(4): p. 924-35.

6. Harada, F. and S. Nishimura, *Possible Anticodon Sequences of $tRNA^{His}$, $tRNA^{Asn}$, and $tRNA^{Asp}$ from Escherichia coli B. Universal Presence of Nucleoside Q in the First Position of the Anticodons of These Transfer Ribonucleic Acids.* Biochemistry, 1972. **11**: p. 301-308.

7. Frey, B., et al., *Mutations in the Escherichia coli fnr and tgt genes - Control of molybdate reductase activity and the cytochrome-D complex fy FNR.* Journal of Bacteriology, 1989. **171**(3): p. 1524-1530.

8. Peden, J., *CodonW.* 1997.

9. Fahlman, R.P., T. Dale, and O.C. Uhlenbeck, *Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites.* Molecular Cell, 2004. **16**(5): p. 799-805.

10. Meier, F., et al., *Queuosine modification of the wobble base in $tRNA^{His}$ influences 'in vivo' decoding properties.* EMBO J., 1985. **4**: p. 823-827.

11. Morris, R.C., K.G. Brown, and M.S. Elliott, *The effect of queuosine on tRNA structure and function.* Journal of Biomolecular Structure & Dynamics, 1999. **16**(4): p. 757-774.

12. Reisser, T., W. Langgut, and H. Kersten, *The nutrient factor queuine protects HeLa cells from hypoxic stress and improves metabolic adaptation to oxygen availability.* Eur J Biochem, 1994. **221**(3): p. 979-986.

13. Hueck, C.J., *Type III protein secretion systems in bacterial pathogens of animals and plants.* Microbiology and Molecular Biology Reviews, 1998. **62**(2): p. 379-+.

14. Durand, J.M.B., et al., *Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of Shigella flexneri: the expression of the virF gene.* Molecular Microbiology, 2000. **35**(4): p. 924-935.

15. Buckingham, R.H., *Codon context and protein synthesis - Enhancements of the genetic code.* Biochimie, 1994. **76**(5): p. 351-354.

16. Waas, W.F., et al., *Role of a tRNA base modification and its precursors in frameshifting in eukaryotes.* Journal of Biological Chemistry, 2007. **282**(36): p. 26026-26034.

17. Garcia, G.A., J.K. Hurt, and Y.-C. Chen, *RNA Modifying Enzymes*, in *Carbohydrates, Nucleosides and Nucleic Acids*, C.-H. Wong and P.G. Wang, Editors. 2009, Elsevier Ltd. p. In Press.

18. Bienz, M. and E. Kubli, *Wild-type $tRNA^{Tyr}$ reads the TMV RNA stop codon, but Q base-modified $tRNA^{Tyr}$ does not.* Nature, 1981. **294**: p. 188-190.

19. Zaher, H.S. and R. Green, *Fidelity at the Molecular Level: Lessons from Protein Synthesis.* Cell, 2009. **136**(4): p. 746-762.

CHAPTER III

Site-specific Modification of *Shigella flexneri virF* mRNA by tRNA-guanine Transglycosylase *in vitro*

Modified nucleosides have been shown to play a role in RNA structure and stability [1-3]. The incorporation of modified nucleosides has been characterized more fully for tRNA (and some other RNAs, *e.g.,* rRNA & snRNA), than for mRNA [4-6]. Mandal and Breaker have characterized binding of small molecules and metabolites to RNA motifs termed “riboswitches,” which are often located in the untranslated regions (UTR) of target RNAs in both prokaryotic and eukaryotic species [7-9]. It has been shown that binding of the small molecule changes the conformation of the RNA and thereby modulates either transcription or translation of the target RNA [10].

Breaker and colleagues have reported the discovery of a riboswitch that responds to the queuine precursor, $preQ_1$, and appears to regulate the expression of the four genes that are involved in $preQ_1$ biosynthesis [11-13]. Durand and colleagues have demonstrated a positive correlation between *virF* mRNA translation and the presence of active TGT (and presumably queuine modification) [14]. It certainly seems possible that base modification of an mRNA could modulate a similar conformational change observed with the $preQ_1$ riboswitch. It therefore is feasible that such a control mechanism for gene

expression might be involved in regulating the expression of virulence factors in pathogenic organisms, as is apparently seen with VirF.

To date, the only known function of TGT is to catalyze the modification of tRNA with queuine; however, previous work has shown that the eubacterial TGT will recognize a U-G-U sequence in alternative RNA structures (*i.e.,* minihelical stem-loop, dimeric cognate tRNA) [15-17]. It is conceivable that an mRNA may also be modified directly by TGT provided the mRNA contained the appropriate recognition elements. As a first step towards probing the possibility that TGT may modulate the translation of VirF via modification of the *virF* mRNA, Michaelis-Menten kinetic analyses were conducted to probe this modification by TGT *in vitro*.

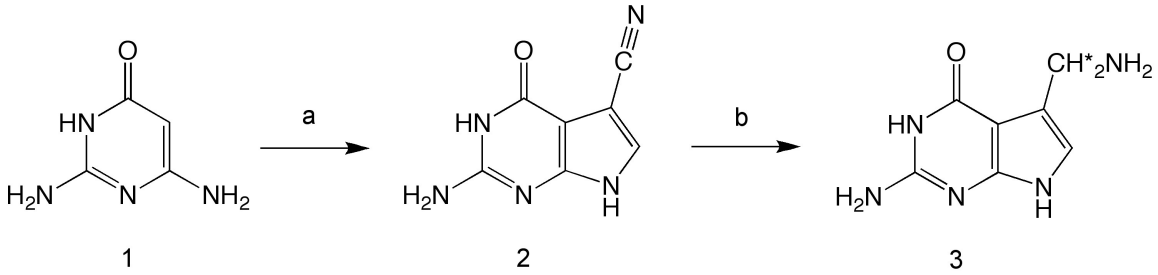
Materials and Methods

Reagents

Unless otherwise specified, all reagents were from Sigma or Aldrich. DNA oligonucleotides, agarose, dithiothreitol (DTT), T4 DNA ligase, and DNA ladders were from Invitrogen. All restriction enzymes and Vent® DNA polymerase were from New England Biolabs. The ribonucleic acid triphosphates (NTPs) and pyrophosphatase were from Roche Applied Sciences. The deoxyribonucleic acid triphosphates (dNTPs) were from Promega. Low-melting Seaplaque agarose was from Cambrex. Gelase™ Enzyme Prep, MasterAmp™ High Fidelity RT-PCR Kit, and Scriptguard™ RNase Inhibitor were from Epicentre. Epicurian coli® XL2-Blue ultracompetent cells were from Stratagene. Amicon Ultra Centrifugal Filter Devices were from Millipore. Whatman GF/C Glass Microfibre Filters and

all bacterial media components were from Fisher. The QIAPrep® Spin Miniprep Kit was from Qiagen. Tris-HCl Buffer was from Acros Organics. [^{3}H]-PreQ$_1$ was obtained from a custom synthesis by American Radiolabeled Chemicals, Inc. T7 RNA polymerase was isolated from *E. coli* BL21 cells containing the plasmid pBH161 according to the procedure of Prof. William McCallister, State University of New York, Brooklyn. The *E. coli* TGT was isolated with an amino terminal his-tag as previously described [18].

Synthesis of [^{3}H] preQ$_1$



1 2 3

Figure III-1. Synthesis of Radiolabeled [^{3}H] PreQ$_1$. *a*) chloro(formyl)acetonitrile, NaOAc, H_2O, 100°C; *b*) $^3H_2$, Pd/C, MeOH, 50 psi. Note that aqueous workup exchanges any tritium in the amine.

The cyano precursor (preQ$_0$, **2**) was synthesized according to the method of Migawa and Townsend by the condensation of chloro(formyl)acetonitrile and pyrimidine **1** [19]. Reduction of the cyano precursor with tritium gas gave the desired radiolabeled substrate preQ$_1$ (**3**) with a specific activity of 500mCi/mmol [20]. After aqueous workup, the tritium in the amine is exchanged out. The tritium reduction was performed commercially by American Radiolabeled Compounds Co.

Construction of pTZvirF

The plasmid pBDG302, containing the *virf* gene, was received from Prof. Glenn Björk (Umeå University, Sweden). The *virf* gene was amplified from the plasmid by polymerase chain reaction (PCR) under the following conditions: pBDG302 (500 ng), PCR primers (20 pmol each, see pTZvirF in Table III-1), Mg^{2+} (2 mM), dNTP mix (1 mM each NTP), Vent DNA polymerase (4 U), brought to a final volume of 50 μL with deionized water. The sample was treated with 30 PCR cycles of the following sequence: 94°C for 1 min., 50°C for 1 min., and 72°C for 2 min., followed by a final extension at 72°C for 5 min. Following a double restriction enzyme digestion with *Pst*I and *Eco*RI (40 U each, 20 μL reaction) for 1 h at 37°C, the PCR product and vector were gel-purified from Seaplaque agarose with Gelase™ according to the vendor's protocol. The purified *virf* gene was then ligated into digested pTZ19R (5:1 volume ratio, 20 μL reaction) following overnight incubation with T4 DNA ligase (2 U) at 17°C. The ligated sample (10 μL) was transformed into 100 μL of Epicurian coli® XL2-Blue ultracompetent cells according to the Stratagene protocol. Cells were grown overnight at 37°C on L-Amp plates (50 μg/mL ampicillin). Individual colonies were isolated, and 3-mL 2xTY (16 g bactotryptone, 10 g yeast extract, 5 g NaCl/liter of water with 50 μg/mL ampicillin) liquid cultures were inoculated at 37°C with shaking. Plasmid was isolated via miniprep, and the *virf* gene sequence was confirmed with DNA sequencing (University of Michigan DNA Sequencing Core Facilities).

Table III-1. Oligonucleotide Sequences for pTZvirF. Key restriction enzyme sites or nucleotides of interest are underlined.

Sample Name	Primer Name	Primer Sequence (5' – 3')
pTZvirF	*Pst*I FWD	GACTGCAGATGATGGATATGGGAC
	*Eco*RI REV	GAGAATTCTTAAAATTTTTTATG
pTZvirF($G_{421}A$) Quik-Change	G421A FWD	GTTTAGCTTGCCTTTTATCAGCTAT-TTCTGATGAGGAAGCTTTATATAC
	G421A REV	GTATATAAAGCTTCCTCATCAGAAA-TAGCTGATAAAAGGCAAGCTAAAC

In vitro Transcription

In vitro transcription reactions with pTZvirF were conducted by first linearizing the plasmid at the end of the *virf* sequence with the restriction enzyme *Eco*RI (40 U/100 µL DNA, 500 µL reaction). The sample was ethanol precipitated at -20°C, and the pellet was resuspended in 250 µL of deionized water. *In vitro* transcription conditions were as follows (1 mL total volume): pTZvirF template (100 µL), transcription buffer (4 mM Tris-HCl, pH 8.0; 2 mM $MgCl_2$, 0.5 mM DTT, 0.1 mM spermidine), NTP mix (4 mM each NTP), T7 RNA polymerase (2500 U), inorganic pyrophosphatase (2 U), and RNase inhibitor (200 U). The reaction was incubated at 37°C for approximately 4 h, and stored at -20°C following transcription. Best results were obtained when the 1-mL reaction was prepared and redistributed into 100 µL volumes prior to incubation at 37°C. The MasterAmp™ High Fidelity RT-PCR Kit was used according to vendor's protocol to generate *virf* DNA, which was confirmed via sequence analysis of the DNA product (Figure III-3).

Mfold Analysis and Synthesis of virF MH

Analysis of the energetically favorable secondary structures within the *virF* mRNA sequence was performed using the biophysical web tool Mfold [21]. Sequences of approximately 10 nucleotides surrounding either side of the six possible recognition motifs were analyzed by the web tool, and the hairpin structure determined to be most favorable was found between nucleotides 410-433 in the *virF* mRNA, which contains the potential recognition sequence $U_{420}G_{421}U_{422}$ (Figure III-2). The RNAture Oligonucleotide Analyzer web tool was used to predict a Tm of 71 °C for the *virF* minihelix hairpin. An Expedite™ Nucleic Acid Synthesis System was used to synthesize the 24-nucleotide sequence (5'-GGAGUAGUCUUUGUCGACUAUUUU-3') using the vendor's protocols for the synthesis of RNA at the 1 µmole scale. The reagents were from Perkin-Elmer and the RNA amidites were from Glen Research. The extinction coefficient determined for this RNA minihelix was ε_{260}=265.3 OD/mM.

Characterization of TGT-virF mRNA Intermediate

Various concentrations of *virF* mRNA (1-20 µM) and *E. coli* TGT (500 nM-20 µM) were incubated at 37°C overnight in transcription buffer supplemented with RNase inhibitor (40 U). A previously characterized minihelical substrate, ECYMH (the anticodon stem-loop of *E. coli* $tRNA^{Tyr}$), was studied as a positive control [16]. Samples were analyzed by SDS-PAGE under mild denaturing conditions as previously described [22]. Product *virF* mRNA containing an abasic site(s) in place of the substrate guanine base were analyzed by RT-PCR as previously described or by chemical cleavage following treatment with

spermidine and/or hydroxylamine. Briefly, samples containing *virF* mRNA and *E. coli* TGT (5 μM each) were incubated in HEPES reaction buffer (100 mM HEPES, pH 7.3; 20 mM $MgCl_2$; 5 mM DTT) in a total volume of 60 μL for 30 min. at 30°C. Samples were treated with either hydroxylamine (420 mM), spermidine (4.2 μM), or a combination of the two for 30 min. at 37°C. The samples were characterized on an ethidium bromide stained, 1.2% formaldehyde agarose gel.

Generation of virF mRNA($G_{421}A$)

The single nucleotide mutation of guanine 421 to adenine in the *virF* mRNA sequence was generated via QuikChange site-directed mutagenesis (Stratagene), producing the new vector pTZvirF($G_{421}A$). The reaction conditions were as follows: pTZvirF(wt) template (800 ng), Quik-Change PCR primers (10 pmol each, see pTZvirF($G_{421}A$) in Table III-1), dNTP mix (0.25 mM each NTP), and Vent DNA polymerase (2 U), brought to a final volume of 30 μL with deionized water. The sample was treated with 25 PCR cycles of the following sequence: 94°C for 30 sec., 50°C for 1 min., and 72°C for 6.5 min. The PCR product was treated for 2 h at 37°C with *Dpn*I (40 U), and addition of NEBuffer 4 (20 mM Tris-Acetate (pH 7.9), 50 mM $K(CH_3COO)$, 10 mM $Mg(CH_3COO)_2$, 1 mM DTT) was required for proper digestion of wildtype plasmid. The digested sample (10 μL) was transformed into 100 μL of XL2-Blue ultracompetent cells according to vendor's protocol (Stratagene). Cells were grown overnight at 37°C on L-Amp plates (50 μg/mL ampicillin). Individual colonies were isolated, and 3 mL 2xTY liquid cultures (with 50 μg/mL ampicillin) were inoculated and incubated at 37°C with shaking. Plasmid was isolated via miniprep (Qiagen), and the *virF*

mRNA($G_{421}A$) mutation was confirmed by DNA sequencing (University of Michigan DNA Sequencing Core Facilities).

Kinetic assays

Assays were conducted by monitoring the incorporation of radiolabeled substrate, [^{3}H]-preQ$_1$, into *E. coli* tRNA^{Tyr}, dG$_{34}$ECYMH (5′-GGGAGCAGACUdGUAAAUCUGCUCCC-3′), and various *virF* mRNA substrates. Samples generated by *in vitro* transcription were concentrated with Amicon Ultra Centrifugal Filters (10,000 MWCO). The concentration of *virF* mRNA was determined with a Cary UV-Visible Spectrophotometer, where the approximate concentration of a single stranded RNA sample A_{260}=1 OD is 40 µg/mL and the molecular weight of *virF* mRNA is 252g/mmol *virF* mRNA. In brief, kinetic assays were set up under the following conditions: RNA substrate (various concentrations), [^{3}H]-preQ$_1$ (10 μM, 296 mCi/mmol stock), TGT (100 or 200 nM, as specified), and HEPES reaction buffer to a final volume of 400 μL. All samples were incubated at 37°C for 5 min. to allow for temperature equilibration. Reactions were initiated by addition of TGT. Aliquots (70 μL) were removed at various points throughout the reaction and quenched in 3 mL of 5% TCA for 1 h before filtering on glass-fiber filters. Each filter was washed with three volumes of 5% TCA and a final wash of ethanol to dry the filter. The samples were analyzed in a scintillation counter (Beckman) for radioactive decay, where counts were reported in DPM and later converted to pmol [^{3}H]-preQ$_1$ by the following conversion (pmol = DPM x 0.00152, for the [^{3}H]-preQ$_1$ stock with a specific activity of 296 mCi/mmol). Initial velocities were determined

by converting the slopes of these plots (pmol/min) to units of sec^{-1}, taking into account the concentration of the enzyme and aliquot size. The individual data points from each trial were averaged, and the standard deviation was determined for each concentration of RNA substrate. The average data points (with error bars representing their standard deviations) were plotted. However, all of the individual data points were fit via nonlinear regression to the Michaelis-Menten equation and the line for that fit is displayed (Figure III-4). All non-linear regression fits with the Michaelis-Menten equation were determined using Kaleidagraph (Abelbeck Software).

For direct incorporation of queuine, *virF* mRNA was also incubated with the human TGT catalytic subunit (hTGTc). Briefly, reactions containing *virF* mRNA (10 μM), hTGTc (1 μM), [^{3}H]-queuine (20 μM), RNase inhibitor (200 U), and bicine buffer in a total volume of 400 μL were incubated at 37°C for 25 min. Aliquots (70 μL) were removed every 5 min, and the samples were prepared as described previously.

Results

Construction and In vitro Transcription of pTZvirF

To provide micromolar quantities of *virF* mRNA for our studies, we generated an *in vitro* transcription clone for the *virF* mRNA. The *virf* gene was subcloned from the plasmid pBDG302 containing the *virf* gene (a gift from Professor Glenn Björk, Umeå University, Sweden) into a plasmid suitable for *in vitro* transcription, generating pTZvirF. *VirF* mRNA was synthesized via *in vitro* run-off transcription following digestion with *Eco*RI, linearizing pTZvirF at the end

of the *virf* gene sequence. The *virF* mRNA was physically characterized on an ethidium bromide-stained, 1.2% formaldehyde agarose gel. Reverse transcriptase / polymerase chain reaction (RT-PCR) was utilized to generate dsDNA from the *in vitro* transcription product using the same oligonucleotide primers initially designed for subcloning of the *virf* gene. Examples of TAE agarose gels of the *virF* mRNA and the dsDNA from the RT-PCR reaction are shown in Figure III-3.

VirF

```
aug aug gau aug gga cau aaa aac aaa aua gau aua aag guu cgc uug cau aac uau auu auu uua uau gca
M   M   D   M   G   H   K   N   K   I   D   I   K   V   R   L   H   N   Y   I   I   L   Y   A

aaa agg UGU uca aug acg guu agc uca ggc aau gaa acu uug acu auc gau gaa ggg caa auu gcu uuu aua
K   R   C   S   M   T   V   S   S   G   N   E   T   L   T   I   D   E   G   Q   I   A   F   I

gag cga aau aua caa aua aac guc ucc aua aaa aaa ucu gau agc auu aau cca uuu gag auu aua agc cuu
E   R   N   I   Q   I   N   V   S   I   K   K   S   D   S   I   N   P   F   E   I   I   S   L

gac aga aau uua uua uua agc auu auu aga aua aug gaa cca auu uau uca uuu caa cac ucc uau ucu gag
D   R   N   L   L   L   S   I   I   R   I   M   E   P   I   Y   S   F   Q   H   S   Y   S   E

gag aaa agg ggg uua aac aaa aaa aua uuc cuc cuc ucu gag gag gag guu ucu auc gau uUG Uuc aaa ucu
E   K   R   G   L   N   K   K   I   F   L   L   S   E   E   E   V   S   I   D   L   F   K   S

aua aaa gag aug ccu uuc ggc aaa aga aag auc uau agu uua gcu ugc cuu uua uca gcU GUu ucu gau gag
I   K   E   M   P   F   G   K   R   K   I   Y   S   L   A   C   L   L   S   A   V   S   D   E

gaa gcu uua uau acu ucg aua ucg aua gcu ucu ucu cuu agu uuu ucu gau cag aua agg aag auU GUu gaa
E   A   L   Y   T   S   I   S   I   A   S   S   L   S   F   S   D   Q   I   R   K   I   V   E

aaa aac auc gag aag aga ugg cgu cuu ucu gau auu uca aau aac uug aau uua uca gaa aua gcU GUu aga
K   N   I   E   K   R   W   R   L   S   D   I   S   N   N   L   N   L   S   E   I   A   V   R

aaa cga uug gag agu gaa aaa uua aca uuu caa caa auc cuu cuu gau auu cgc aug cau cau gca gca aag
K   R   L   E   S   E   K   L   T   F   Q   Q   I   L   L   D   I   R   M   H   H   A   A   K

cuu uua uug aau agu caa agc uau auu aau gaU GUa uca aga cuu auc gga aua uca agc cca ucu uau uuu
L   L   L   N   S   Q   S   Y   I   N   D   V   S   R   L   I   G   I   S   S   P   S   Y   F

aua agg aaa uuu aau gaa uau uau ggu aua acu cca aag aaa uuu uac uua uau cau aaa aaa uuu uaa
I   R   K   F   N   E   Y   Y   G   I   T   P   K   K   F   Y   L   Y   H   K   K   F   *
```

Figure III-2. Nucleotide Sequence of *virF* mRNA. Six UGU sequences are found in the sequence, and are highlighted in red.

The single-stranded *virF* mRNA appears to run on the gel at approximately 500 bp in comparison to the double-stranded DNA ladder. The size of the *virf* gene is 789 bp, and the corresponding mRNA is 789 nucleotides in length. We hypothesize that the mRNA is running at a lower “apparent” molecular weight

due to the propensity of mRNA to adopt a variety of conformations, even in an agarose gel. This would explain why the observed molecular weight is a little larger than one half the size of the double-stranded *virf* DNA. The RT-PCR product (dsDNA) migrated to the anticipated molecular weight for a DNA sample of approximately 800 bp. DNA sequencing of this product confirmed the RT product was the *virf* gene.

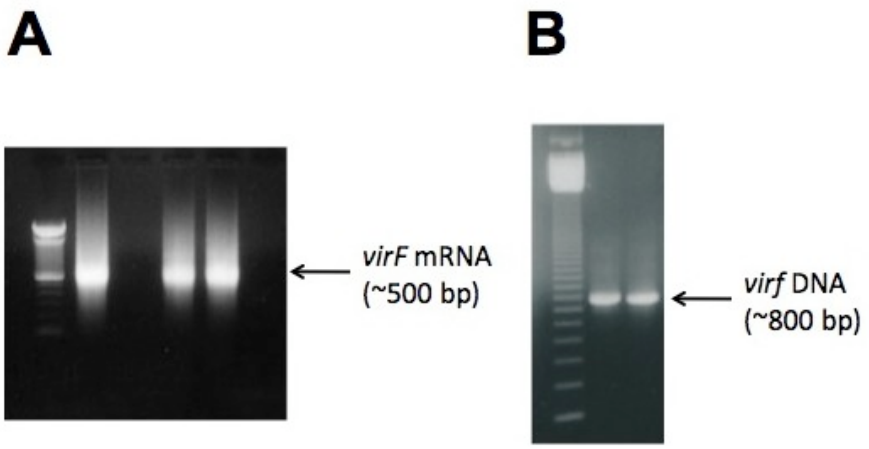


Figure III-3. Transcription of *virF* mRNA. A) *In vitro* transcription of *virF* mRNA visualized with ethidium bromide on a 1% TAE agarose gel. The single-stranded mRNA ran at approximately 500 bp in comparison to the 100 bp DNA ladder (lane 1). Lane 2: *virF* mRNA from a previous transcription reaction. Lanes 4 and 5: *virF* mRNA from replicate transcription reactions. B) RT-PCR samples visualized with ethidium bromide on a 1% TAE agarose gel. RT-PCR was conducted using extracted mRNA from replicate samples (lanes 2 & 3), electrophoresed in comparison to a 100 bp DNA ladder (lane 1). Each RNA sample produced the corresponding 800 bp DNA species indicative of the *virf* ORF.

Kinetic Analysis of E. coli tRNA^{Tyr}, ECYMH Minihelix with preQ$_1$

For comparison, Michaelis-Menten kinetic analyses were conducted with the natural RNA substrate *E. coli* tRNA^{Tyr} (ECY) and the modified minihelix substrate dG$_{34}$ECYMH (the anticodon stem-loop of E. coli tRNA^{Tyr} where the

guanosine at position 34 contains a 2′-deoxyribose) with [^{3}H] preQ$_1$. It has been shown previously that a minihelix RNA consisting of the anticodon arm and loop of a queuine-cognate tRNA is a sufficient substrate for TGT [15]. Aliquots were taken at various time points over a 15-minute incubation of 100 nM *E. coli* TGT, various concentrations of ECY (0.05-1.5 μM) or dG$_{34}$ECYMH (0.05-5 μM), and saturating concentrations of [^{3}H]-preQ$_1$. The kinetic constants determined for the incorporation of [^{3}H]-preQ$_1$ with ECY and dG$_{34}$ECYMH are shown in comparison with the kinetic data for the *virF* substrates in Table III-2.

Attempt at Characterization of TGT-virF mRNA Covalent Intermediate

TGT and *virF* mRNA were incubated for an increased period of time (overnight) in an attempt to visualize the formation of a covalent intermediate, which has been characterized previously with both tRNA and minihelical substrate RNA [22]. Visualization of the complex by SDS-PAGE and Coomassie Blue staining under mild denaturing conditions was unsuccessful. Only the 45 kDa unbound *E. coli* TGT was detected (data not shown). In addition, chemical cleavage of the product mRNA with hydroxylamine/spermidine lead to degradation of the RNA species. Due to issues with RNA integrity, the technique was unsuccessful in determining the site of queuine modification.

Kinetic Analysis of virF mRNA, virF MH Minihelix with preQ$_1$

Using the same approach described above, Michaelis-Menten kinetic analyses were conducted with *virF* mRNA. Aliquots were taken at various time points over a 1 h incubation of 200 nM TGT, various concentrations of *virF* mRNA (0.1-10 μM), and saturating concentrations of [^{3}H]-preQ$_1$ (Figure III-4A).

Higher concentrations of *virF* mRNA were tested to obtain an accurate kinetic profile by characterizing the reaction over a large range of concentrations. In addition to characterizing the wildtype *virF* mRNA, a *virF* minihelix RNA (*virF* MH) corresponding to the 410-433 hairpin sequence (underlined in Figure III-2, structure in Figure III-4B) as well as a full-length *virF* mRNA mutant ($G_{421}A$) were studied. The kinetic analyses were performed with 100 nM *E. coli* TGT, various concentrations of *virF* MH (0.1-10 μM), and saturating concentrations of [^{3}H]-preQ$_1$ (Figure III-4B). The full-length *virF* mRNA($G_{421}A$) was incubated under the same conditions as the wildtype mRNA, but only at concentrations corresponding to K_M and 5xK_M, as determined from the kinetic constants of *virF* mRNA(wt) (Table III-2). A "no RNA" control was also included to determine the background level of radioactivity present in the samples (Figure III-4C). The data in Figures III-4A & B were fit by non-linear regression to the Michaelis-Menten equation and those in Figure III-4C by linear regression.

Both the full-length *virF* mRNA(wt) and the *virF* minihelix exhibited RNA concentration-dependent incorporation of [^{3}H]-preQ$_1$ over time, and the Michaelis-Menten equation provided a good fit for the data. The full-length *virF* mRNA($G_{421}A$), which is full-length mRNA with a single nucleotide mutation at guanine 421, was analyzed at both 2 μM (~ K_M) and 10 μM (~ 5xK_M) mRNA. The *virF* mRNA($G_{421}A$) mutant showed no detectable activity greater than the "no RNA" control (Figure III-4C). The kinetic constants determined for the *virF* mRNA substrates are shown in Table III-2. Both the *virF* mRNA and *virF* MH have K_M values in the low micromolar range, even though k_{cat} and k_{cat}/K_M for both

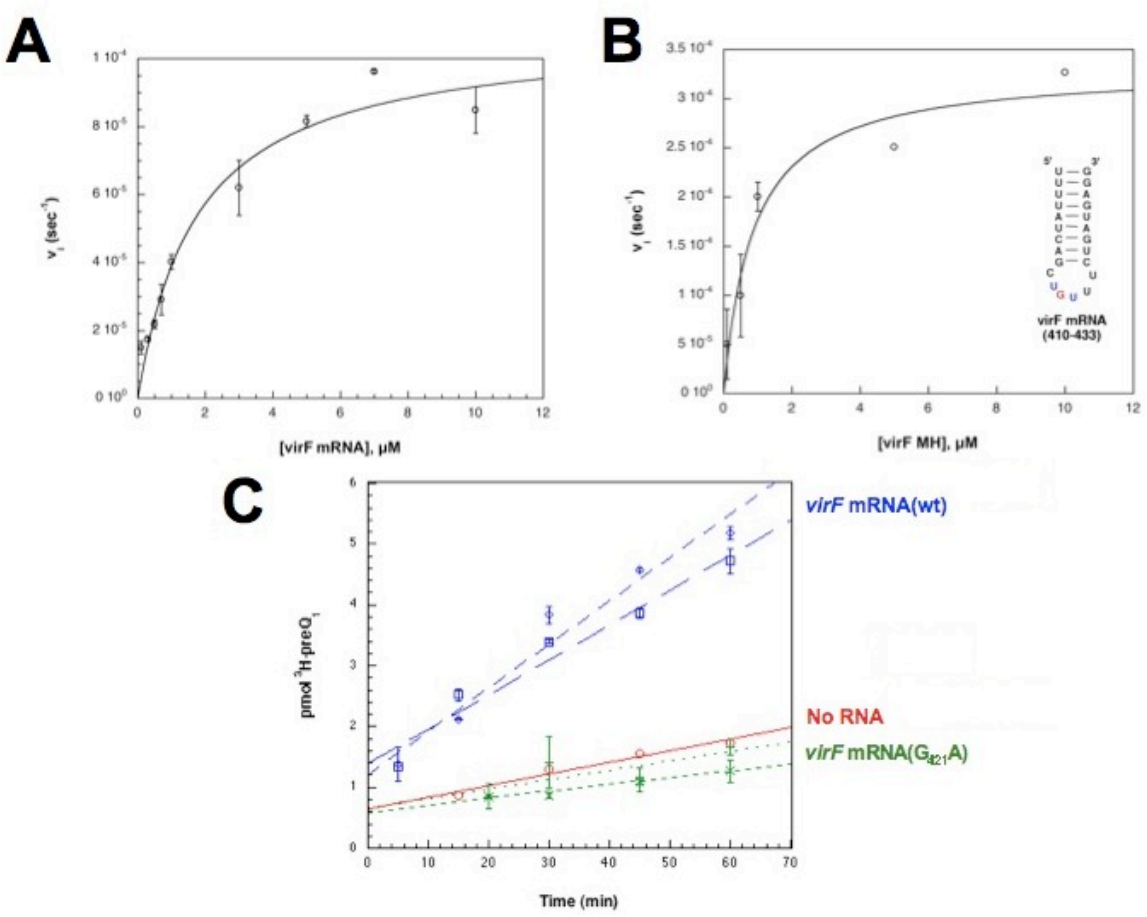


Figure III-4. Kinetic Characterization of *virF* mRNA Substrates with *E. coli* TGT and [^{3}H] PreQ$_1$. A) and B) The individual data points were fit to the Michaelis-Menten equation by non-linear regression, and the average data points (with error bars representing the standard deviations of the averages) for each concentration of RNA substrate are depicted. The study was performed with duplicate kinetic trials at saturating concentrations of [^{3}H] preQ$_1$. The inset in B shows the structure of the RNA minihelix corresponding to bases 410-433 of the *virF* mRNA. C) The comparison of radioactive substrate incorporated over time in *virF* mRNA(wt) (blue), *virF* mRNA(G$_{421}$A) (green), and a no RNA control (red), fit by linear regression. Each *virF* mRNA substrate was analyzed at two concentrations: 2 µM and 10 µM. The average of duplicate trials is shown.

is lower than the corresponding values for the ECY substrates. Attempts to modify *virF* mRNA with queuine using human hTGTc were unsuccessful. We

now know that this is most likely due to an inactive preparation of human TGT (Chen & Garcia, unpublished).

Table III-2. RNA Kinetic Parameters with *E. coli* htTGT(wt) and [^{3}H] PreQ$_1$ at pH 7.3.

	K_M (μM)	k_{cat} (10^{-4} s^{-1})	k_{cat}/K_M (10^2 M^{-1} s^{-1})
ECY	0.16 (0.03)[a]	40 (2)	250 (40)
dG$_{34}$ECYMH	0.26 (0.07)	70 (5)	270 (60)
virF mRNA	1.8 (0.4)	1.1 (0.1)	0.60 (0.1)
virF MH	0.87 (0.3)	3.3 (0.3)	3.8 (1)

[a] Standard errors are in parentheses.

Discussion

VirF is a critical transcriptional regulator responsible for activating virulence genes in *Shigella flexneri*. Durand and colleagues demonstrated the involvement of tRNA-guanine transglycosylase (TGT) in modulating the translation of VirF via the observation that a mutant strain of *Shigella* with an inactivated *tgt* gene (termed *vacC)* showed decreased virulence [23]. VirF protein levels were dramatically lower in the mutant as compared to wildtype, but the *virF* mRNA levels showed no detectable difference from wildtype *S. flexneri*. The lack of VirF resulted in a reduction of all downstream virulence gene expression, and thus exhibited a less virulent phenotype than that of the wildtype bacterium. When transformed with a plasmid encoding the *S. flexneri tgt* gene, both VirF translation and virulence were restored in the *vacC* mutant. It had previously been shown that the presence of modified nucleosides enhances translation [24, 25]. However, other studies have shown that growth rate and protein translation as a whole are not directly affected by a lack of queuine-

modified tRNA [26]. While this interesting correlation between VirF translation and TGT activity has been known for some time, the exact role TGT plays in the translation of this primary virulence factor remains unclear.

Our laboratory has previously demonstrated that TGT can modify substrates with more unusual structures than a canonical tRNA fold. We reported that a dimeric form of the *E. coli* $tRNA^{Tyr}$ serves as a substrate for TGT, with a slightly higher K_M and identical k_{cat}, relative to the normal tRNA [17]. It had previously been shown from NMR studies that the anticodon arms of the dimer subunits were intact and pointing away from the center of the dimer [27].

Those studies demonstrate that TGT can recognize a minihelix containing the requisite U-G-U sequence even in the context of a larger RNA structure. TGT is not the first tRNA modification enzyme to demonstrate recognition of alternative RNA structures. Gu and coworkers have shown that, *in vitro*, the modification enzyme tRNA (m^5U54)-methyltransferase will methylate 16S ribosomal RNA from *E. coli* in addition to its physiological tRNA substrate [5], although they found no evidence for this occurring *in vivo*. The enzyme pseudouridine synthase catalyzes the isomerization of specific uridines to the modified nucleoside ψ. A single pseudouridine synthase has been shown to have “dual specificity”, recognizing and modifying both tRNA and snRNA [28-30]. Ofengand and colleagues have also characterized a critical pseudouridine synthase responsible for site-specific modification with pseudouridine *in vivo* for both tRNA and 23S rRNA in *E. coli* [31, 32]. Such precedence for tRNA modification enzymes to recognize and modify other RNA species *in vitro* and *in*

vivo, suggests that *virF* mRNA modification mediated by TGT may also occur. As a first step to probe for this possibility, we examined the *virF* mRNA sequence for the presence of a U-G-U sequence in a TGT recognition motif. Of the six U-G-U sequences in the *virF* mRNA, the one involving bases 410-433 (Figure III-2 and 4B) was predicted by Mfold analysis to be able to fold into a hairpin structure possibly suitable for recognition by TGT.

Incubation of *virF* mRNA with *E. coli* TGT and radiolabeled $preQ_1$ revealed that the *virF* mRNA is indeed a substrate for TGT *in vitro*. The kinetic parameters (K_M, k_{cat}, Table III-2) for *virF* mRNA were determined from a nonlinear fit of the data to the Michaelis-Menten equation (Figure III-4A). The *virF* mRNA substrates exhibit the same trend in kinetic parameters as the ECY substrates, where the minihelix substrates (dG_{34}ECYMH and *virF* MH) have slightly higher values for both K_M and k_{cat}/K_M with respect to the corresponding full-length RNA substrates (Table III-2). The values of K_M for both ECY and *virF* mRNA are very similar, both in the low micromolar range. This is encouraging considering the size and structural difference between the two substrates, where ECY is approximately 80 nucleotides in length with a very well-defined tertiary structure common to tRNA and the *virF* mRNA is ca. 800 nucleotides in length and presumably does not have a compact tertiary structure. The kinetic data for the *virF* mRNA (Figure III-4A) fits to a single K_M, consistent with a single site of modification within the *virF* mRNA. The k_{cat} for the *virF* mRNA is ca. 40-fold lower than that for tRNA. It has previously been shown that altering the position of the U-G-U sequence within the minihelix loop of cognate tRNAs is correlated

with a reduction in activity [33]. All of the biochemical and structural data previously reported is consistent with a covalent intermediate, via Asp264, in the TGT reaction [18, 34-36]. The hairpin loop of the ECY substrate contains 7 nucleotides, whereas the hairpin in the *virF* mRNA substrate contains only 6 nucleotides in the loop. This difference in loop length may result in a suboptimal orientation of the guanosine ribose in the U_{420}-G_{421}-U_{422} loop in the *virF* mRNA, making nucleophilic attack by Asp264 somewhat less likely. This difference in loop length and orientation could account for the reduced k_{cat} that we have observed for the *virF* mRNA substrates.

Our analysis of the *virF* mRNA sequence predicts that there should be a single site of modification, guanine 421. We have taken two approaches to investigate this. In our first approach, an RNA minihelix (*virF* MH) corresponding to the predicted hairpin structure within the native *virF* mRNA sequence (bases 410-433) was chemically synthesized (Figure III-4B). The stem consists of nine base pairs, where the first four nucleotides in the stem are uridine residues forming wobble pair interactions with three guanines and one Watson-Crick pair with an adenosine. At first glance, the stability of a minihelix with three G-U wobble pairs might be questionable; however, in the context of the *virF* mRNA, the ends of the helix may be held in close proximity by other intramolecular interactions. Additionally, the *virF* MH by itself has a predicted melting temperature of 71ºC, indicating the structure should be stable at physiological and assay temperatures. The *virF* MH is a substrate for TGT *in vitro*. The K_M values for both the full-length *virF* mRNA and *virF* MH (1.8 μM and 0.87 μM,

respectively; Table III-2) are very similar, suggesting that the minihelix structure is likely a predominant conformation in the *virF* mRNA.

Although the recognition of the *virF* minihelix by TGT is consistent with the *virF* mRNA serving as a substrate for TGT, it does not provide conclusive evidence that guanine 421 is the site of modification in the *virF* mRNA. There are six UGU sequences within the *virF* mRNA that are possible recognition sites for TGT. Therefore, our second approach was to construct the point mutation, $G_{421}A$, in the full-length *virF* mRNA to demonstrate the importance of guanine 421. The *virF* mRNA($G_{421}A$) mutant resulted in a complete loss of activity at two different concentrations of RNA (Figure III-4C), indicating that G_{421} is indeed essential for recognition by TGT. Had a second exchangeable guanine existed in the sequence, we would have expected to see a decreased or possibly even unchanged activity of the mRNA. The relationship between the "no RNA" negative control and the *virF* mRNA($G_{421}A$) indicates that guanine 421 is the only exchangeable nucleotide in the *virF* mRNA sequence (at least within the concentration ranges tested), and that the kinetic parameters observed for *virF* mRNA(wt) are due to specific recognition by TGT and could not be attributed to non-specific interactions with this large nucleic acid molecule. It should be noted that, under the conditions of the assay (Figure III-4C), it appears that the enzyme is undergoing a limited number of turnovers. Two factors may be contributing to this. The first is that our calculations of kinetic parameters assume 100% active enzyme, which is almost certainly an over-estimate. The active enzyme concentration may be as much as two-fold lower as recent studies suggest that

the eubacterial TGT may exist as a homodimer with “half-of-sites” reactivity [37]. This would effectively double the turnovers per active site. Second, the off-rate for the modified mRNA may be sufficiently slower (relative to that for tRNA) such that the turnover rate may indeed be significantly slowed under these conditions. It remains to be seen if these observations hold under *in vivo* conditions.

Conclusions

The *E. coli* TGT, which has 99% sequence identity to the *S. flexneri* TGT, does indeed recognize the *virF* mRNA as a substrate *in vitro*. Further, we show that this recognition results in the site-specific modification of a single base in the *virF* mRNA (G_{421}) with mutagenesis and *virF* mRNA minihelix studies. With a K_M value in the low micromolar range, it is very possible that the modification of *virF* mRNA may be biologically relevant (*e.g.*, may occur *in vivo*). These results provide the first “proof of principle” evidence that post-transcriptional RNA modification may regulate mRNA function, as it has long been recognized to do for tRNA. The recent work characterizing the $preQ_1$ riboswitch revealed that it is a fairly simple hairpin structure (the simplest riboswitch structure characterized to date) [11]. Such a simple structure could feasibly occur in the coding region of an mRNA species. In fact, the TGT modification site that we have discovered in the *virF* mRNA is predicted to occur in a simple hairpin structural motif. Base modification in a hairpin structure in the coding sequence of *virF* mRNA could induce a similar structural switch as seen in the $preQ_1$ riboswitch and thereby influence translation of VirF.

Notes to Chapter III

This chapter has been published in Hurt, J. K., Olgen, S., Garcia, G. A. (2007). Site-specific modification of *Shigella flexneri virF* mRNA by tRNA-guanine transglycosylase *in vitro*, *Nucleic Acids Research*, **35**(14): 4905-13. We gratefully acknowledge Professor Glenn Björk for the plasmid pBDG302 containing the *virf* gene. It is worth noting that ongoing research in the Garcia lab suggests the active human TGT functions as a heterodimer. We are currently working to characterize the activity of this eukaryl enzyme. It may be worth revisiting the queuine modification with *virF* mRNA and *in vitro* translation experiments with this optimized protocol.

Abbreviations used: TGT, tRNA-guanine transglycosylase; UTR, untranslated region; $preQ_1$, 7-aminomethyl 7-deazaguanine; DTT, dithiothreitol; NTP, ribonucleotide triphosphate; dNTP, deoxyribonucleotide triphosphate; RT-PCR, reverse transcription polymerase chain reaction; PCR, polymerase chain reaction; Tm, melting temperature; ECY, *E. coli* $tRNA^{Tyr}$; ECYMH, *E. coli* $tRNA^{Tyr}$ minihelix; MH, minihelix; HEPES, 4-(2-hydroxyethyl-1-piperazineethanesulfonic acid); TCA, trichloroacetic acid; DPM, disintegrations per minute; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; NMR, nuclear magnetic resonance; snRNA, small nuclear RNA.

1. Agris, P.F., *The importance of being modified: Roles of modified nucleosides and Mg2+ in RNA structure and function*, in *Progress in Nucleic Acid Research and Molecular Biology, Vol 53*, W.E. Cohn and K. Moldave, Editors. 1996, Academic Press Inc: 525 B Street, Suite 1900, San Diego, CA 92101-4495. p. 79-129.

2. Varani, G. and I. Tinoco, *RNA Structure And NMR-Spectroscopy.* Quarterly Reviews Of Biophysics, 1991. **24**(4): p. 479.

3. Heus, H.A. and A. Pardi, *Structural Features That Give Rise To The Unusual Stability Of Rna Hairpins Containing Gnra Loops.* Science, 1991. **253**(5016): p. 191.

4. Grosjean, H., M. Sprinzl, and S. Steinberg, *Posttranscriptionally modified nucleosides in transfer RNA: Their locations and frequencies.* Biochimie, 1995. **77**(1-2): p. 139-141.

5. Gu, X.G., J. Ofengand, and D.V. Santi, *In Vitro Methylation of Escherichia coli 16S rRNA by tRNA (m^5U54)-Methyltransferase.* Biochemistry, 1994. **33**(8): p. 2255-2261.

6. Noon, K.R., E. Bruenger, and J.A. McCloskey, *Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile Sulfolobus solfataricus.* J Bacteriol, 1998. **180**(11): p. 2883-2888.

7. Barrick, J.E., et al., *New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control.* Proceedings Of The National Academy Of Sciences Of The United States Of America, 2004. **101**(17): p. 6421.

8. Mandal, M. and R.R. Breaker, *Gene regulation by riboswitches.* Nature Reviews Molecular Cell Biology, 2004. **5**(6): p. 451.

9. Sudarsan, N., J.E. Barrick, and R.R. Breaker, *Metabolite-binding RNA domains are present in the genes of eukaryotes.* RNA-A Publication Of The RNA Society, 2003. **9**(6): p. 644.

10. Coppins, R.L., K.B. Hall, and E.A. Groisman, *The intricate world of riboswitches.* Current Opinion in Microbiology, 2007. **10**(2): p. 176-181.

11. Roth, A., et al., *A riboswitch selective for the queuosine precursor preQ(1) contains an unusually small aptamer domain.* Nature Structural & Molecular Biology, 2007. **14**(4): p. 308-317.

12. Kim, J.N. and R.R. Breaker, *Purine sensing by riboswitches.* Biology of the Cell, 2008. **100**(1): p. 1-11.

13. Meyer, M.M., et al., *Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria.* RNA-A Publication of the RNA Society, 2008. **14**(4): p. 685-95.

14. Durand, J.M.B., et al., *Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of Shigella flexneri: the expression of the virF gene.* Molecular Microbiology, 2000. **35**(4): p. 924-935.

15. Curnow, A.W., et al., *tRNA-Guanine Transglycosylase from Escherichia coli: Gross tRNA Structural Requirements for Recognition.* Biochemistry, 1993. **32**: p. 5239-5246.

16. Curnow, A.W. and G.A. Garcia, *tRNA-Guanine Transglycosylase from Escherichia coli - Minimal tRNA Structure and Sequence Requirements for Recognition.* Journal of Biological Chemistry, 1995. **270**(29): p. 17264-17267.

17. Curnow, A.W. and G.A. Garcia, *tRNA-Guanine Transglycosylase from Escherichia coli: Recognition of Dimeric, Unmodified tRNA^{Tyr}.* Biochimie, 1994. **76**(12): p. 1183-1191.

18. Kittendorf, J.D., et al., *tRNA-guanine transglycosylase from Escherichia coli: Molecular mechanism and role of aspartate 89.* Biochemistry, 2001. **40**(47): p. 14123-14133.

19. Migawa, M.T., et al., *A Two Step Synthesis of the Nucleoside Q Precursor 2-Amino-5-cyanopyrrolo[2,3-d]pyrimidin-4-one (PreQ$_0$).* Synthetic Communications, 1996. **26**(17): p. 3317-3322.

20. Cheng, C.S., et al., *Synthesis of pyrrolo[2,3-d]pyrimidines that are structurally related to methylated guanosines from tRNA and the nucleoside Q analogs, PreQ(0) and PreQ(1).* Nucleos Nucleot, 1997. **16**(4): p. 347-364.

21. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Research, 2003. **31**(13): p. 3406-3415.

22. Chervin, S.M., J.D. Kittendorf, and G.A. Garcia, *Probing the Intermediacy of Covalent RNA-Enzyme Complexes in RNA Modification Enzymes.* Methods in Enzymology, 2007. **425**: p. 121-137.

23. Durand, J.M., et al., *vacC, a Virulence-associated Chromosomal Locus of Shigella flexneri, is Homologous to tgt, a Gene Encoding tRNA-Guanine Transglycosylase (TGT) of Escherichia coli K-12.* Journal of Bacteriology, 1994. **176**(15): p. 4627-4634.

24. Björk, G.R., *Genetic Dissection of Synthesis and Function of Modified Nucleosides in Bacterial Transfer RNA.* Progress in Nucleic Acids Resarch and Molecular Biology, 1995. **50**: p. 263-338.

25. Qian, Q.A., J.F. Curran, and G.R. Bjork, *The methyl group of the N-6-methyl-N-6-threonylcarbamoyladenosine in tRNA of Escherichia coli modestly improves the efficiency of the tRNA.* J Bacteriol, 1998. **180**(7): p. 1808-1813.

26. Noguchi, S., et al., *Isolation and Characterization of an Escherichia coli Mutant Lacking tRNA-Guanine Transglycosylase.* Journal of Biological Chemistry, 1982. **257**(11): p. 6544-6550.

27. Rordorff, B.F. and D.R. Kearns, *Nuclear Magnetic Resonance Investigation of the Base-Pairing Structure of Escherichia coli tRNA^{Tyr} Monomer and Dimer Conformations.* Biochemistry, 1976. **15**(15): p. 3320-3330.

28. Massenet, S., et al., *Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase Pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA.* Molecular & Cellular Biology, 1999. **19**(3): p. 2142-2154.

29. Maden, B.E.H., *The numerous modified nucleotides in eukaryotic ribosomal RNA.* Progress in Nucleic Acids Research and Molecular Biology, 1990. **39**: p. 241-300.

30. Gu, X.R., et al., *Molecular recognition of tRNA by tRNA pseudouridine 55 synthase.* Biochemistry, 1998. **37**(1): p. 339-343.

31. Horne, D.A., et al., *RNA substrate recognition by E-coli pseudouridine synthase (rluA) with dual substrate specificity.* Abstracts Of Papers Of The American Chemical Society, 1997. **213**: p. 308.

32. Raychaudhuri, S., et al., *Functional effect of deletion and mutation of the Escherichia coli ribosomal RNA and tRNA pseudouridine synthase RluA.* Journal Of Biological Chemistry, 1999. **274**(27): p. 18880.

33. Nonekowski, S.T. and G.A. Garcia, *tRNA Recognition by the E. coli TGT: the Role of U33 in U-G-U Sequence Recognition.* RNA, 2001. **7**(10): p. 1432-1441.

34. Goodenough-Lashua, D.M. and G.A. Garcia, *tRNA-Guanine Transglycosylase from Escherichia coli: a Ping-Pong Kinetic Mechanism is Consistent with Nucleophilic Catalysis.* Bioorganic Chemistry, 2003. **31**(4): p. 331-344.

35. Kittendorf, J.D., et al., *An essential role for aspartate 264 in catalysis by tRNA-guanine transglycosylase from Escherichia coli.* Journal of Biological Chemistry, 2003. **278**(43): p. 42369-42376.

36. Xie, W., X.J. Liu, and R.H. Huang, *Chemical trapping and crystal structure of a catalytic tRNA guanine transglycosylase covalent intermediate.* Nature Structural Biology, 2003. **10**(10): p. 781-788.

37. Stengl, B., et al., *Crystal structures of tRNA-guanine transglycosylase (TGT) in complex with novel and potent inhibitors unravel pronounced induced-fit adaptations and suggest dimer formation upon substrate binding.* Journal of Molecular Biology, 2007. **370**(3): p. 492-511.

CHAPTER IV

Troubleshooting Expression and Purification of Recombinant VirF

VirF (30 kDa, 263 residues) is the master transcriptional activator of the virulence cascade in *Shigella spp.* Due to the importance of this virulence factor in the infection cycle of the bacterium, biochemical characterization of VirF has been of interest in many research groups. The importance of VirF was first determined by genetic knockout experiments in *Shigella flexneri* [1]. Further analyses performed *in vivo* in *S. flexneri* allowed for identification of primary downstream targets, namely VirB and VirG(IcsA) [2-4]. The mechanisms that regulate VirF expression have also been extensively studied [5-10]. Of particular interest is the temperature-sensitive repression mechanism of the *virf* promoter by H-NS, which also represses many other virulence gene promoters. Although much knowledge has been gained regarding the virulence cascade and mechanisms of *S. flexneri* infection, relatively little is known about the VirF protein itself.

In vitro characterization of VirF was first reported by Tobe *et al.* [11]. Though expression of recombinant VirF in *E. coli* failed, the authors characterized the relevant DNA-binding sites in the *virB* promoter using a maltose-binding protein (MBP) fusion with VirF, MalE-VirF. Transcriptional activation of *virB* by MalE-VirF was studied, in addition to repression by H-NS, to probe the molecular basis of the dual-regulation mechanism of *S. flexneri*

virulence. Porter and Dorman also attempted expression of recombinant VirF *in vivo* [12]. The authors generated various point mutants of VirF that were expressed as soluble protein, by both site-directed mutagenesis and random mutagenesis with nitrosoguanidine. While monitoring transcriptional activation in a mutant *S. flexneri* (Δ*virf*) strain, the authors monitored activity of the mutant VirF proteins by activation (through VirB) of a *mxiC-lacZ* fusion to identify residues important for structure and function.

Details about the structure of VirF have been inferred from the crystal structures of other AraC family members [13, 14]. With the many technological advances in molecular biology, there are a variety of protocols that aim to improve expression of "difficult" proteins (*i.e.,* fusion tags, molecular chaperones). Expression and purification of wildtype VirF would allow for further characterization of the virulence factor, especially the N-terminal dimerization domain, which remains poorly understood. In this study, we aim to employ new techniques to express recombinant VirF.

Materials and Methods

Reagents

Unless otherwise specified, all reagents were from Sigma Aldrich. Gelase™ Agarose-Gel Digesting Preparation was from Epicentre. The QIAprep Spin Miniprep Kit was from Qiagen. Bactotryptone and yeast extract were from Fisher Scientific. Chemically competent BL21(DE3) cells, Ni^{+2}-NTA resin, Lysonase™ Bioprocessing Reagent, and Factor Xa Cleavage Capture kit were from Novagen. Amicon Ultra-4 Centrifugal Filter Units were from Millipore.

Ultracompetent XL-2 Blue cells were from Stratagene. Plasmids harboring molecular chaperones (GroEL, GroES, DnaK, DnaJ, GrpE) were from Takara. All restriction enzymes, Vent® DNA polymerase, calf-intestinal alkaline phosphatase (CIP), amylose resin, and pMAL-c2x were from New England Biolabs. SeaPlaque agarose was from Cambrex. Isopropyl β-D-thiogalactoside (IPTG), T4 DNA Ligase, V5-HRP antibody, Positope™ Control Protein, pBAD202-D/TOPO, LMG194 chemically competent cells, Expressway™ Cell-free Expression System and all synthetic oligonucleotides were from Invitrogen. The ActivePro Kit was from Ambion. HisProbe-HRP and CN/DAB substrate kit was from Pierce. L-[^{35}S]-methionine was purchased from Perkin-Elmer. GF-C glass fiber filters were from Whatman. The deoxynucleotide triphosphates (dNTPs, monosodium salts) were from Promega. Polyacrylamide gels, SDS-PAGE buffer strips, low molecular weight protein standard, and all FPLC columns were from GE Healthcare. The nitrocellulose membranes and blot paper were from BioRad.

Bacterial Stains and Plasmids

All *E. coli* strains (with the exception of XL-2 Blue Ultracompetent cells) were purchased as glycerol stocks, and chemically competent cells were prepared according to vendor protocol. The *virf* gene in the housekeeping vector pTZ19R (pTZvirF) was used for all subsequent subcloning reactions (See Chapter III for details).

Constructs for Expression of Full-length Wildtype VirF

Lactose-inducible Expression Plasmids

Using pTZvirF as a template (see Chapter III for details), the *virf* gene was subcloned into the dual promoter vector, pLacT7 [15]. In a total volume of 50 µL, samples containing 1 µg pTZvirF, PCR primers (10 pmol each, see pLacT7-virF in Table IV-1), dNTP mix (2.5 mM each NTP), 2 mM $MgSO_4$, and Vent® DNA polymerase (2 U) were incubated according to the following temperature sequence: 30 cycles – 94°C for 1 min., 65°C for 1 min., and 72°C for 2 min. The vector pLacT7 and the *virf* PCR product (20 µL DNA) were treated with *Nde*I and *Bam*HI in a double restriction enzyme digestion (40 U each, 30 µL volume) at 37°C for 90 min. The vector and PCR product were gel-purified from Seaplaque agarose gel with Gelase™ according to the vendor's protocol. In a total volume of 20 µL, the *virf* PCR product and digested pLacT7 (5:1 volume ratio) were incubated overnight at 17°C with T4 DNA ligase (2 U). The ligated sample (pLacT7-virF, 10 µL) was transformed into 100 µL Ultracompetent XL-2 Blue cells according to the vendor's protocol and plated on L-Amp plates (50 µg/mL ampicillin). Individual colonies were isolated, and grown in either 3 mL 2xTY or M9 cultures with 50 µg/mL ampicillin at 37°C with shaking (M9 Media, per 1L dH_2O: 1 mM $MgSO_4$, 0.2% glycerol, 0.00005% thiamine, 0.1% casamino acids and M9 salts (6 g Na_2HPO_4, 3 g KH_2PO_4, 1 g NH_4Cl, 500 mg NaCl, 3 mg $CaCl_2$). Plasmid was isolated via miniprep, and the correct sequence was confirmed by DNA sequencing (University of Michigan DNA Sequencing Core Facilities).

To monitor VirF expression from pTZvirF directly, a dsDNA cassette was designed for insertion of the *E. coli* consensus ribosome-binding site (RBS, AGGAGG) between the T7 promoter and the *virf* gene (see pTZvirF+RBS in

Table IV-1). The palindromic DNA sequence was designed with the restriction enzyme site *Pst*I at both the 5' and 3' ends. The vector pTZvirF and dsDNA cassette (10 µL DNA) were treated with *Pst*I in a double restriction enzyme digestion (20 U, 20 µL volume) at 37°C for 1h. The vector and insert were gel-purified from Seaplaque agarose gel with Gelase™ according to vendor protocol. The purified pTZvirF sample (100 µL) was treated with alkaline phosphatase (CIP, 30 U) in NEBuffer 3 (50 mM Tris-HCl (pH 7.9), 100 mM NaCL, 10 mM $MgCl_2$, 1 mM DTT) at 37°C for 1h. The reaction was quenched by phenol-chloroform extraction (1:1 ratio, buffered phenol (pH 8.0) and 24:1 chloroform:isoamyl alcohol). The aqueous layer was removed to a new microcentrifuge tube and precipitated for 1h at -20°C with 0.1 volume 3 M NaOAc (pH 5.3) and 2.5 volumes 100% EtOH. In a total volume of 20 µL, the RBS cassette and digested pTZvirF (5:1 volume ratio) were incubated overnight at 17°C with T4 DNA ligase (2 U). The ligated sample (pTZvirF+RBS, 10 µL) was transformed into 100 µL Ultracompetent XL-2 Blue cells according to the vendor's protocol and grown on L-Amp plates (50 µg/mL ampicillin) at 37°C. Individual colonies were isolated and sequenced as previously described.

Arabinose-inducible Expression Plasmid

pBADvirF was constructed following amplification of the *virf* gene for insertion into the directional pBAD202 vector via TOPO cloning. In a total volume of 50 µL (performed in triplicate), samples containing 1 µg pTZvirF, PCR primers (10 pmol each, see pBADvirF in Table IV-1), dNTP mix (2.5 mM each NTP), and Vent® DNA polymerase (2 U) were incubated according to the

Table IV-1: Oligonucleotide Sequences for VirF Expression Plasmid Construction. Key restriction enzyme sites or nucleotides of interest are underlined.

Sample Name	Primer Name	Primer Sequence (5' – 3')
pTZvirF+RBS	VirF Insert RBS	CTGCAGGAGGTACCTCCTGCAG
pTZvirF+RBS +V5	Add V5 FWD	CTGCAGGAGGTACGTACATATGGGTAAGCC-TATCCCTAACCCTCTCCTCGGTCTCGATTC-TACGCTGCAG
	Add V5 REV	CTGCAGCGTAGAATCGAGACCGAGGAGAG-GGTTAGGGATAGGCTTACCCATATGTACGT-ACCTCCTGCAG
pET20b-virF	pET FL FWD	GAGCCATGGATGATGGATATG
	BamHI REV	*See pLacT7-virF BamHI REV
pET20b-virF(29-263)	pET M29 FWD	GACCATGGATGACGGTTAGCTC
	BamHI REV	*See pLacT7-virF BamHI REV
pLacT7-virF	NdeI FWD	GCATGCCTGCATATGATGGATATGGAC
	BamHI REV	CGACGGCCAGTGGATCCTTAAAATTTTTTATG
pBADvirF	pBAD FWD	CACCATGGCTTAATACGACTCACTATAGGG
	BamHI REV	*See pLacT7-virF BamHI REV
pBADvirF +His	pBAD FWD	*See pBADvirF pBAD FWD
	Add His REV	GAGGGATCCGTAAAATTTTTTATGATATAAG
pMALvirF	XmnI FWD	CTGAAATCGTTCGGGACATAAAAAC
	EcoRI REV	GAGAATTCTTAAAATTTTTTATG
pMALvirF Quik-Change	QC (-G) FWD	CTCGGGATCGAGGGAAGGCTGAAATCTTC-GGGACATAAAAACAAAATAG
	QC (-G) REV	CTATTTTGTTTTTATGTCCCGAAGATTTCAG-CCTTCCCTCGATCCCGAG

following temperature sequence: 30 cycles – 94°C for 1 min., 55°C for 1 min., and 72°C for 2 min. All replicate reactions were combined and ethanol precipitated at -20°C overnight. The precipitate was collected by centrifugation, and the DNA pellet was resuspended in water (15 µL). In a total volume of 6 µL,

2.5 ng *virf* PCR product and pBAD202-D/TOPO were ligated and transformed into chemically competent OneShot® TOP10 cells according to the vendor's protocol. Samples were grown on L-Kan plates (50 µg/mL kanamycin) at 37°C, and colonies were isolated as previously described.

In a 200 µL total volume, purified pBADvirF (100 µL DNA) was treated with *Nco*I (70 U) at 37°C for 1h to remove upstream affinity tags (~376 bp). Samples were incubated at 65°C for 20 min. to inactivate the restriction enzyme, and the pBADvirF sample (50 µL digested DNA) was incubated overnight with T4 DNA ligase (2 U, 57 µL total volume) at 17°C. The re-ligated plasmid (10 µL) was transformed into 100 µL Ultracompetent XL-2 Blue cells according to vendor protocol and grown at 37°C on L-Kan plates. The new vector is termed the corrected pBADvirF.

Constructs for Expression of Full-length VirF with Fusion Tags

Histidine (His) Tag

Constructs for lactose-inducible expression of full-length VirF were prepared in pET vectors from Novagen. The same *virf* PCR product prepared for insertion into pLacT7 above was subcloned into pET28a (yielding pET28a-virF with an N-terminal 6x His-tag) following the same protocol. In collaboration with Prof. Oleg Tsodikov and Dr. Tappan Biswas, the *virf* gene was also subcloned into the expression vector pET19b (with N-terminal 10x His-tag) between the *Nde*I and *Xho*I sites. The vector (pET19bpps) was manipulated to incorporate the PreScission protease cleavage site (LEVLFQGP) for removal of the fusion tag following purification. For the arabinose-inducble pBADvirF, the *virf* gene

was amplified from pTZvirF+RBS with a single base change (stop codon converted to a tyrosine codon) incorporated into the reverse oligonucleotide to allow read-through of a C-terminal 6x His-tag (see pBADvirF+His in Table IV-1). Reactions were set up as described above for pBADvirF with the following exception: the annealing temperature was adjusted to 50°C in the PCR cycles.

V5 Epitope

For insertion of an N-terminal V5 epitope (GKPIPNPLLGLDST) into VirF, a 70 bp dsDNA cassette was synthesized (see pTZvirF+RBS+V5 in Table IV-1). In a 40 µL total volume, the complimentary oligonucleotides (10 µL each) were pre-incubated at 37°C for 5 min. The dsDNA cassette and pTZvirF (10 µL DNA, 20 µL total volume) were treated with *Pst*I (40 U) at 37°C for 1 h. The insert and plasmid were gel-purified and subsequently treated as described for pTZvirF+RBS.

pelB Leader Sequence

The *virf* gene was subcloned into pET20b (Novagen) carrying the N-terminal *pelB* leader sequence (MKYLLPTAAAGLLLLAAQPAMA) for co-translational transport to the periplasm. In a total volume of 50 µL, samples containing 1 µg pTZvirF, PCR primers (10 pmol each, see pET20b-virF in Table IV-1), dNTP mix (1.5 mM each NTP), 2 mM $MgSO_4$, and Vent® DNA polymerase (2 U) were incubated according to the following temperature sequence: 30 cycles – 94°C for 1 min., 50°C for 1 min., and 72°C for 2 min. The vector pET20b (5 µL DNA) and the *virf* PCR product (10 µL DNA) were treated with *Nco*I and *Bam*HI via double restriction digestion (20 U each, 20 µL volume) at 37°C for 2 h. The

vector and PCR product were gel-purified from a Seaplaque agarose gel with Gelase™ according to vendor protocol. In a total volume of 20 µL, the *virf* PCR product and digested pET20b (5:1 volume ratio) were incubated overnight at 17°C with T4 DNA ligase (2 U). The ligated sample (pET20b-virF, 10 µL) was transformed into 100 µL Ultracompetent XL-2 Blue cells according to the vendor's protocol and plated on L-Amp plates (50 µg/mL ampicillin). Individual colonies were isolated as previously described.

malE, Maltose-binding Protein (MBP)

Construction of a MalE-VirF fusion protein has been previously described [11]. In a total volume of 50 µL, samples containing 1 µg pTZvirF, PCR primers (10 pmol each, see pMALvirF in Table IV-1), dNTP mix (0.25 mM each NTP), 2 mM $MgSO_4$, and Vent® DNA polymerase (2 U) were incubated according to the following temperature sequence: 30 cycles – 94°C for 1 min., 50°C for 1 min., and 72°C for 2 min. The vector pMAL-c2x and the *virf* PCR product (3 µL DNA) were treated with *Xmn*I (40 U, 30 µL volume) at 37°C for 1 h. The samples were ethanol precipitated at -20°C overnight as previously described. The DNA pellet was collected by centrifugation and resuspended in water (25 µL). Both samples were treated with *Eco*RI (40 U, 30 µL volume) in a subsequent restriction enzyme digestion at 37°C for 1 h. The vector and PCR product were gel-purified from Seaplaque agarose gel with Gelase™ according to vendor protocol. In a total volume of 20 µL, the *virf* PCR product and digested pMAL-c2x (3:1 volume ratio) were incubated overnight at 17°C with T4 DNA ligase (2 U). The ligated sample (pMALvirF, 10 µL) was transformed into 100 µL Ultracompetent XL-2 Blue cells

according to the vendor's protocol and plated on L-Amp plates (50 µg/mL ampicillin). Individual colonies were isolated as previously described.

The *Xmn*I site was ligated in a blunt-end manner, adding seven base pairs to the start of the *virf* gene. One base was removed (G2697) by site directed mutagenesis according to vendor protocol to restore the proper reading frame (Stratagene). In a total volume of 30 µL, samples containing 500 ng pMALvirF, Quik-Change PCR primers (10 pmol each, see pMALvirF Quik-Change in Table IV-1), dNTP mix (0.25 mM each NTP), 2 mM $MgSO_4$, and Vent® DNA polymerase (2 U) were incubated according to the following temperature sequence: 20 cycles – 94°C for 1 min., 50°C for 1 min., and 72°C for 6.5 min. The amplified samples were treated with *Dpn*I (40 U) in 1x NEBuffer 4 (20 mM Tris-Acetate (pH 7.9), 50 mM $K(CH_3COO)$, 10 mM $Mg(CH_3COO)_2$, 1 mM DTT) at 37°C for 2 h to digest wildtype DNA. The digested sample (10 µL) was transformed into 250 µL chemically competent TG2 cells and grown overnight on L-Amp plates (50 µg/mL ampicillin). Plasmid DNA was isolated, and the sequence was confirmed as previously described.

Constructs for Expression of VirF Fragments

VirF (29-263)

A fragment of the *virf* gene (705 bp) corresponding to a truncation in the first 84 bp was amplified from pTZvirF. The resultant protein product would have an N-terminal deletion of 28 amino acids, but retain a portion of the dimerization domain and the full linker region and DNA-binding domain. This *virf* fragment was subcloned into pET20b according to the protocol described for the full-length

virf gene (see pET20b-virF(29-263) in Table IV-1 for oligonucleotide sequences). In collaboration with Prof. Tsodikov and Dr. Biswas, this fragment was also subcloned into pET19bpps as described previously. The *virf* ($G_{421}A$) mutant of pET19b-virF(29-263) was prepared according to the protocol discussed in Chapter III.

VirF (154-263)

The *virf* fragment (297 bp) corresponding to the 99-amino acid DNA-binding domain was amplified and subcloned into pET19bpps as previously described.

Cell-based Expression Trials

SDS-PAGE Analysis

Denaturing polyacrylamide gels were run with samples diluted 1:1 with SDS-PAGE buffer (50 µL, 60 mM Tris•HCl (pH 6.8), 2% SDS, 10% glycerol, 5% β-mercaptoethanol, 0.01% bromophenol blue). The low molecular weight standard is run as a molecular weight marker (6 major protein markers (kDa): 97, 66, 45, 30, 20, 14). All samples (4 µL) were boiled for 5 min., and loaded on a 8-25% gradient polyacrylamide gel (SDS buffer strips: 8.9 mM Tris base, 8.9 mM boric acid, 0.2 mM EDTA, 2% agarose) run under denaturing conditions according to the PhastGel™ protocol (GE Healthcare). Polyacrylamide gels were visualized by Coomassie Blue staining according to the vendor's protocol unless otherwise specified.

Western Blot Protocol

Samples were run on polyacrylamide gels under denaturing conditions and transferred to a nitrocellulose membrane in ice-cold Towbin buffer (3.03 g Tris base, 14.4 g glycine, 200 mL methanol per 1L dH_2O) at 60 V for 30 min. The membrane was washed with phosphate-buffered saline with Tween (PBST, 0.1 M $NaPO_4$ (pH 7.2), 0.15 M NaCl, 0.05% Tween-20) for 10 min. at room temperature. The horseradish peroxidase (HRP) conjugated primary antibody (HisProbe-HRP or V5-HRP) was added (1:5000) in PBST in a total volume of 10 mL. The membrane was incubated at room temperature for 1 h with shaking. The membrane was washed with a second volume of PBST before addition of the CN/DAB substrate (4-chloro-1-naphthol/3,3-diaminobenzidine, tetrahydrochloride) in Stable Peroxide Substrate Buffer according to the vendor's protocol (Pierce). The Positope™ Control Protein was run in parallel as the positive control for His-tag or V5 epitope expression.

Lactose-inducible Expression Vectors

For expression of full-length untagged VirF, several T7-based vectors (pTZvirF+RBS, pLacT7-virF) were prepared and transformed into BL21(DE3) and K12 (Δ*tgt*, DE3). Individual colonies were isolated and grown at 37°C in 3 mL 2xTY cultures (with appropriate antibiotics). Several fusion-tagged constructs were also prepared to monitor VirF expression in cells. For pTZvirF+RBS+V5, pET28a-virF, pET20b-virF, and pET19b-virF, expression was monitored in both BL21(DE3) and K12 (Δ*tgt*, DE3) cells. Individual colonies were isolated for growth in 3 mL 2xTY media (with appropriate antibiotic) at 37°C overnight. Fresh

media was inoculated with the starter culture (1:100), and samples were tested in parallel with varying concentrations IPTG (0.5-5.0 mM) in 3 mL 2xTY media or with Studier's ZYP-5052 autoinduction media (both supplemented with the appropriate antibiotic) [16]. Following an appropriate induction period (4 hours for IPTG samples and overnight for auto-induction cultures), a small aliquot (100 µL) was removed from each culture along with an uninduced control, and the cells were harvested by centrifugation. The media was discarded, and the cell pellet was resuspended in SDS-PAGE buffer and analyzed by denaturing polyacrylamide gel.

L-arabinose-inducible Expression Vectors

pBADvirF+His was transformed into *E. coli* TG2, LMG194(Δ*ara*) and K12(Δ*tgt*) cells for expression trials. Individual colonies were isolated and grown in 3 mL minimal M9 media (supplemented with 50 µg/mL kanamycin) at 37°C overnight. Fresh M9 minimal media (5 mL, supplemented with 2% L-arabinose and 50 µg/mL kanamycin) was inoculated with the starter culture (1:100) and grown at 37°C with shaking for 4 h. Aliquots (1 mL) were removed from each sample, and the cells were harvested by centrifugation. The supernatant was removed, and each pellet was resuspended in lysis buffer (250 µL, 10 mM Tris•HCl (pH 8.0), 1 mM EDTA, 0.5 mg/mL lysozyme, 0.1 mg/mL DNase I, 10 mM $CaCl_2$). The samples were subjected to three rapid freeze/thaw cycles, frozen on dry ice then placed in a 42°C water bath for 1 min. each. The samples were pelleted at 4°C, and the soluble portion was transferred to a fresh microcentrifuge tube. The soluble fraction was analyzed by SDS-PAGE as a 1:1

mixture of supernatant and SDS-PAGE buffer described previously (50 µL total volume). The insoluble pellet was resuspended in 250 µL SDS-PAGE buffer. Protein species of interest were excised from the polyacrylamide gel and placed in dH_2O for trypsin digestion and subsequent mass spectrometry analysis (University of Michigan Proteome Consortium)

Expression Trials with MalE-VirF

MalE-VirF expression from the *tac*-promoter based vector, pMALvirF, was monitored in BL21 cells. A 500 mL 2xTY culture with 50 µg/mL ampicillin was inoculated with a pMALvirF starter culture (1:100). The culture was incubated at 37°C with shaking to mid-log phase (OD_{600} ≅ 0.6), and MalE-VirF expression was induced upon addition of 5 mM IPTG. Following a 4 h incubation at 37°C, the cells were harvested by centrifugation, and the cell pellet was stored overnight at -20°C. The cells were resuspended in VirF storage buffer (10 mL, 20 mM Tris•HCl (pH 7.4), 0.2 M NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol) supplemented with PMSF in isopropanol (phenylmethylsulfonyl fluoride, 100 mM) and 10 µL Lysonase™ Bioprocessing Reagent (Novagen). The lysis reaction proceeded at room temperature for 15 min., followed by sonication on ice (6 rounds, 10 sec. at maximum setting). The crude extract was purified by amylose affinity chromatography according to the vendor's protocol (New England Biolabs). Column fractions were analyzed by polyacrylamide gel and Coomassie Blue staining as previously described. Samples containing MalE-VirF (70 kDa) were combined and purified by size-exclusion chromatography in VirF storage buffer.

Cleavage of the MalE fusion was performed with Factor Xa (recognition site: I(E/D)GR) according to the vendor's protocol (Novagen). In a total volume of 50 µL, samples containing MalE-VirF (10 µg), and various concentrations of Factor Xa (0.1-0.5 U) were incubated at varying temperatures (4-37°C). Aliquots (10 µL) were removed over a 24 hour incubation period to monitor cleavage of the MBP fusion by polyacrylamide gel with Coomassie Blue staining as previously described. Reactions were performed in the presence and absence of a linear DNA fragment of the *virB* operator or supercoiled pTB601 (1-5 µL DNA) (see Chapter V for details).

Expression Trials with VirF Fragments

Expression of VirF (29-263)

pET19b-virF(29-263) was transformed into chemically competent BL21(DE3) cells according to vendor protocol. A 500 mL ZYP-5052 culture with 50 µg/mL ampicillin was inoculated with starter culture (10 mL), and expression of VirF(29-263) was induced by autoinduction as previously described for the full-length VirF protein. Soluble and insoluble fractions were isolated, and the insoluble pellet was resuspended in 8 M urea. Samples (4 µL) were analyzed by SDS-PAGE and stained with Coomassie Blue or transferred to a nitrocellulose membrane for treatment with HisProbe-HRP in a Western blot procedure according to vendor protocol (Pierce). The denatured His-tagged VirF(29-263) sample was applied to Ni-NTA resin and purified from contaminants by addition of 8 M urea supplemented with 250 mM imidazole. The protein was refolded with step-wise buffer exchange in VirF storage buffer and decreasing concentrations

of urea (8 M, 4 M, 2 M, 0 M) in Amicon Ultra-4 centrifugal filter units at 7,000 x g for 10 min. at 4°C.

Expression of VirF(29-263) was monitored in the presence of various molecular chaperones from Takara. pET19b-virF(29-263) and pET19b-virF(29-263, V141I) were transformed into BL21(DE3) cells containing either pGro7 or pKJE7 (expressing GroEL/ES or DnaK, DnaJ, and GrpE, respectively). Induction of the VirF(29-263) protein was performed according to the same protocol regardless of chaperone identity. A 500 mL 2xTY culture (100 µg/mL carbenicillin, 30 µg/mL chloramphenicol) supplemented with 0.5 mg/mL L-arabinose and inoculated with starter culture (10 mL) was incubated at 37°C with shaking. IPTG (1 mM) was added at mid-log phase, and the culture was incubated at 16°C with shaking overnight. Soluble and insoluble fractions were isolated as previously described. Protein samples were analyzed by SDS-PAGE and Coomassie Blue staining.

Expression of VirF(154-263)

Purification and analysis of the VirF(154-263) protein was performed in collaboration with Prof. Tsodikov and Dr. Biswas. BL21(DE3) cells harboring pET19b-virF(154-263) were induced at OD_{600} = 0.5 with 0.5 mM IPTG at 16 ^{o}C overnight. The protein was purified from the soluble fraction on Ni-IMAC (GE LifeSciences) column followed by heparin chromatography and gel filtration chromatography on an S-200 column (GE LifeSciences). Purified VirF(154-263) protein was incubated with a series of overlapping double-stranded DNA oligomers covering the *virB* promoter region (-131 to -11 in the virB operator)

previously shown to be important for regulation by VirF [11]. The gel mobility shift assay was conducted in VirF reaction buffer (50 mM Tris base (pH 7.0), 100 mM NaCl, 0.05 mM EDTA, 10% glycerol).

In vitro Translation Trials with Full-length VirF

A variety of cell-free *in* vitro translation kits were utilized to monitor expression of the VirF protein. *virF+RBS+V5* mRNA was generated by *in vitro* transcription as previously described [17]. VirF expression was first studied in the Expressway™ cell-free extract (Δ*slyD*) according to vendor protocol (Invitrogen). Various concentrations of *virF+RBS+V5* mRNA (0.3-2.5 μM) were tested. Samples (100 μL) were incubated at 30°C and 37°C for 5 h in parallel. The soluble and insoluble fractions were separated according to the protocol described for pBADvirF induction trials. Samples were loaded on 8-25% gradient polyacrylamide gels and analyzed by Coomassie Blue stain as well as anti-V5 Western blot according to the vendor protocol (BioRad).

Expression of VirF *in vitro* was also monitored in the ActivePro™ Kit *E. coli* S30 extract according to vendor protocol (Ambion). Samples (50 μL) containing either pTZvirF+RBS+V5 or *virF+RBS+V5* mRNA were incubated at 37°C for 2 h, and reactions were quenched with ice-cold acetone (50 μL). Soluble and insoluble fractions were prepared as previously described. All samples (4 μL) were loaded on 8-25% gradient polyacrylamide gels and analyzed by Coomassie Blue staining. T7-based plasmids expressing the human calmodulin-like 3 protein (CALML-3, 19.5 kDa) and chlorampenicol

acetyltransferase (CAT, 25 kDa) were run in parallel as positive controls for *in vitro* translation.

An *E. coli* BL21(DE3) cell-free S12 extract was prepared according to the protocols of Kim *et al*. [18, 19]. A starter culture of BL21(DE3) was used to inoculate 500 mL 2xTY (1:100). The sample was grown at 37°C with shaking to mid-log phase (OD_{600} = ~0.7). IPTG was added to a final concentration of 1 mM, and the culture was induced at 37°C with shaking for 5 h to enhance expression of the T7 RNA polymerase. The cells were harvested by centrifugation (6,000 x g) for 15 min. at 4°C. The supernatant was removed, and the pellet was washed three times with Buffer A (20 mL/g cells, 10 mM Tris-acetate buffer (pH 8.2), 1 mM DTT, 14 mM $Mg(CH_3COO)_2$, 60 mM potassium glutamate, 0.05% β-mercaptoethanol). The pellets were stored at -80°C overnight. The BL21(DE3) cells were resuspended in 10 mL Buffer B (1.3 mL/g cells, 10 mM Tris-acetate buffer (pH 8.2), 1 mM DTT, 14 mM $Mg(CH_3COO)_2$, 60 mM potassium glutamate). Cell lysis was performed by French press (two rounds at 20,000 psi.), and the lysate was collected by centrifugation at 12,000 x g for 10 min. at 4°C. The lysate (8 mL) was incubated at 37°C with shaking for 30 min. in the presence of preincubation solution (2.5 mL, 293 mM Tris-acetate buffer (pH 8.2), 4.4 mM DTT, 2 mM $Mg(CH_3COO)_2$, 10.4 mM ATP (adenosine triphosphate), 100 mM FBP (fructose-1,6-bisphosphate), amino acid mix (0.04 mM each), 23.4 µg/mL pyruvate kinase). The *E. coli* BL21(DE3) S12 extract was stored in individual aliquots (27 µL) at -80°C.

In a total volume of 100 µL, reactions containing test plasmid DNA (0.67 µg/mL), BL21(DE3) S12 extract (27 µL), 57 mM HEPES-KOH buffer (pH 8.2), NTP mix (1.2 mM ATP, 0.85 mM CTP, GTP, and UTP), 2 mM DTT, 90 mM potassium glutamate, 80 mM $NH_4(CH_3COO)$, 12 mM $Mg(CH_3COO)_2$, folinic acid (34 µg/mL), amino acid mix (2 mM each, -Met), 2% PEG-8000 (polyethylene glycol), 33 mM FBP, pyruvate kinase (5 µg/mL), RNase inhibitor (2 U), and L-[^{35}S]-methionine (0.33 mCi/mL) were incubated at 37°C. Aliquots (10 µL) were removed every 2 h over the 6 h incubation period. Samples treated 37°C 1 M NaOH (300 µL) for 10 min. to deacylate any charged tRNAs. The reactions were quenched with 2.5 mL ice-cold 5% TCA (trichloroacetic acid), vortexed briefly, and incubated on ice for 5 min. Samples were filtered on GF-C glass fiber filters and the radioactivity was measured by scintillation counting as previously described [17]. The negative control for radioactivity incorporation was monitored from samples without DNA, and the positive control for protein synthesis was measured in samples containing the CAT expression plasmid. Various VirF expression plasmids were tested (*i.e.,* pTZvirF+RBS+V5, pET19b-virF (full-length, 29-263, and 154-263), pMALvirF). Reactions were performed in the presence and absence of a linear DNA fragment of the *virB* operator or supercoiled pTB601 (1 µL DNA) (see Chapter V for details).

Results

Expression Trials with Full-length VirF

The original *virf* vectors used for cloning and *in vitro* transcription (pTZvirF and pTZvirF($G_{421}A$), see Chapter III for details) were not designed for protein

expression due to the absence of a ribosome-binding site. To monitor VirF expression in cells, a variety of plasmids were constructed to express untagged and fusion-tagged forms of the transcriptional activator. Cell growth in the presence of untagged VirF expression (pTZvirF+RBS, pLacT7, pBADvirF) on solid media resulted in mucoidy colonies. In the case of pTZvirF+RBS (BL21(DE3)) and pBADvirF (LMG194), the mucoid growth was diffuse over the agar plate, with satellite colonies protruding into the solid media. With pLacT7 (BL21(DE3)), cell growth was displayed as "burst colonies," with a single, large colony surrounded by very small, salt-like colonies. When isolated for growth at 37°C in liquid culture, all samples exhibited dense, mucoidy growth attached to the sterile toothpick used for colony selection, while the surrounding media was essentially clear. Due to growth issues, these constructs were not carried forward in cell-based expression trials.

Various solubility (fusion) tags were employed with the hopes of increasing the concentration of soluble VirF in cells. *E. coli* harboring the 6x His-tagged construct, pET28a, exhibited the same mucoidy phenotype in solid and liquid media described previously. While there was some mucoidy nature in minimal M9 media, LMG194 cells transformed with pBADvirF+His exhibited more uniform growth than was previously observed. In induction trials with L-arabinose (2% weight to volume), a band was observed by SDS-PAGE with Coomassie staining at approximately 30 kDa in both the soluble and insoluble fractions (Figure IV-1).

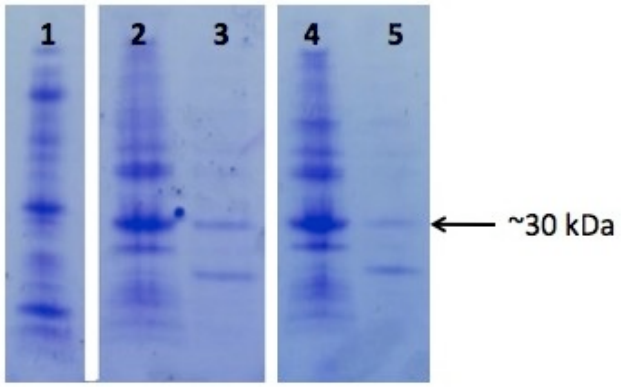


Figure IV-1: Expression of Full-length VirF. Expression of VirF in *E. coli* TG2 and K12 (Δ*tgt*) from pBADvirF monitored by SDS-PAGE and Coomassie Blue staining. Lane 1, low molecular weight standard; lane 2, TG2 insoluble fraction; lane 3, TG2 soluble fraction; lane 4, K12 (Δ*tgt*) insoluble fraction; lane 5, K12 (Δ*tgt*) soluble fraction.

Both samples were excised from the polyacrylamide gel and submitted to the University of Michigan Proteome Consortium for MS/MS analysis following trypsin digestion. Sequence analysis of the resulting peptide fragments confirmed the major identity of the protein sample was the kanamycin resistance protein. It was observed that a low concentration of VirF was also present. Expression of 10x His-tagged VirF (pET19bpps-virF in BL21(DE3)) resulted in normal cell growth in liquid media, yet no protein was expressed as measured by Coomassie staining. Since cytosolic over-expression of full-length VirF may be resulting in misfolding/aggregation, the *virf* gene was subcloned into pET20b with an N-terminal *pelB* fusion. The leader sequence directs the protein to the periplasm, and has been shown to assist in protein folding [20]. BL21(DE3) cells harboring pET20b-virF demonstrated normal cell growth in 2xTY, but again, no soluble VirF was detected by Coomassie stain (results not shown).

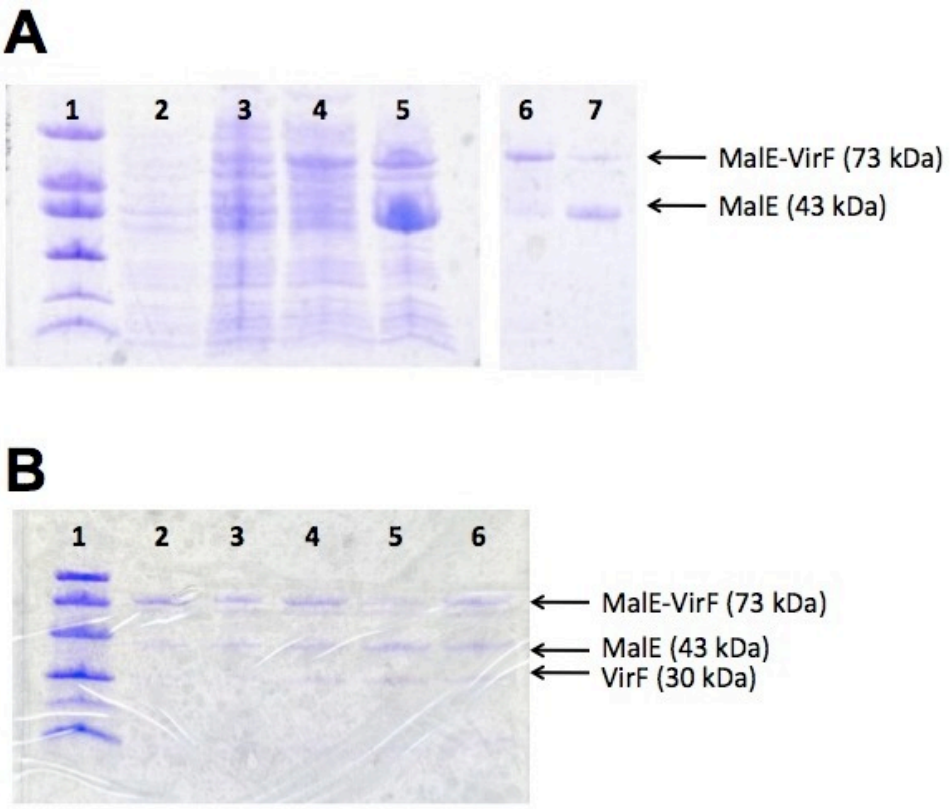


Figure IV-2: Expression and Purification of MalE-VirF. A) MalE-VirF was expressed in BL21 cells and purified by amylose affinity resin and size-exclusion chromatography. Protein was analyzed by SDS-PAGE and Coomassie Blue staining. Lane 1, low molecular weight standard; lane 2, uninduced culture; lane 3, induced with 5 mM IPTG; lane 4, crude extract; lane 5, eluate following amylose resin purification; lane 6, purified MalE-VirF; lane 7, major contaminating protein. B) Representative Factor Xa cleavage reaction at room temperature with 0.5 U Factor Xa (lanes 2-5) or 0.2 U Factor Xa (lane 6). Lane 1, low molecular weight standard; lane 2, 3 h reaction time; lane 3, 5 h reaction time; lane 4, 7 h reaction time; lane 5, 24 h reaction time; lane 6, 24 h reaction time.

Expression of soluble VirF in the form of a maltose-binding protein (MBP) fusion was previously reported [11]. pMALvirF was constructed according to the same protocol, using pMAL-c2x for cytosolic expression of the fusion protein. The vector was also designed with a Factor Xa recognition site between the *malE* and *virf* genes, allowing for cleavage of the fusion protein. Expression in *E.*

coli BL21(DE3) cells induced with 5 mM IPTG yielded soluble protein (Figure IV-2A). This 70 kDa fusion protein, MalE-VirF, was purified by amylose resin and eluted with 10 mM maltose, which competitively binds the maltose-binding protein and displaces the polypeptide from the resin. The major contaminating protein (likely endogenous MBP, 43 kDa) was separated by size-exclusion chromatography.

Due to presumed increased difficulty in X-ray structural characterization of the fusion protein, MalE-VirF was treated with Factor Xa to obtain native VirF. Several different reaction conditions to enhance the stability of VirF in the Factor Xa cleavage reaction were tested (*i.e.*, varying the enzyme to protein ratio, length of reaction time, temperature, available DNA binding site). Analysis by SDS-PAGE at various time points in the reaction followed a disappearance of the 70 kDa fusion protein and the appearance of the 43 kDa MalE. At lower reaction temperatures, a small concentration of the 30 kDa VirF protein was detected after a 24 h incubation period (Figure IV-2B), but this species could not be isolated. In an attempt to improve the stability of VirF, purified pTB601, which contains the virB promoter sequence, or a PCR-amplified dsDNA fragment corresponding to the virB promoter were added to the reactions. None of the above trials were successful in isolating native VirF. This phenomenon of VirF instability in the Factor Xa reaction was previously observed [11].

Expression Trials of Truncated Forms of VirF

In a collaborative effort with Prof. Oleg Tsodikov and Dr. Tappan Biswas, a truncated form of the VirF protein (beginning with methionine 29) was

expressed. The protein product (~ 31 kDa with 10x His-tag) was initially expressed in BL21(DE3) cells and monitored by Coomassie staining and Western blot analysis. The total protein sample illustrated a marked level of VirF(29-263) overexpression (Figure IV-3A). Isolation of the soluble and insoluble fractions confirmed that the VirF(29-263) protein is insoluble.

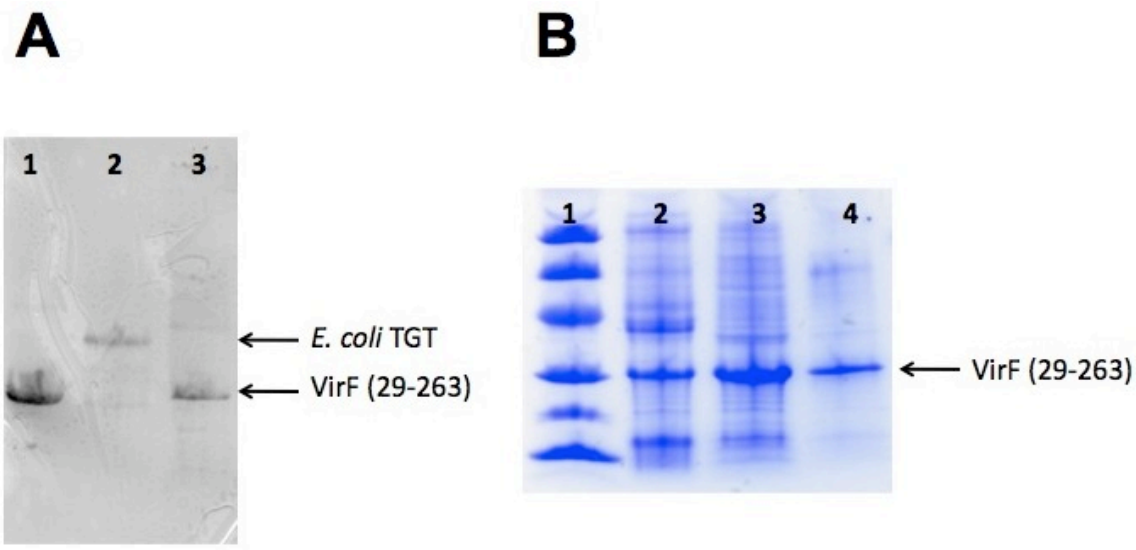


Figure IV-3: Expression of VirF(29-263). A) Anti-His Western blot of total protein sample measuring VirF(29-263) expression in *E. coli* BL21 cells. Lane 1, low molecular weight standard; lane 2, *E.coli* TGT; lane 3, VirF(29-263). B) Purification of denatured VirF(29-263) by Ni-NTA affinity chromatography analyzed by SDS-PAGE and Coomassie Blue staining. Lane 1, low molecular weight standard; lane 2, flow-through; lanes 3-4, elute fractions.

Purification under native conditions was unsuccessful due to the low concentration of soluble VirF(29-263). The insoluble protein was resuspended in 8 M urea to denature the inclusion bodies and attempt purification under denaturing conditions. The denatured protein sample was applied to a nickel column at room temperature, and was purified almost to homogeneity after several rounds of application and elution with 8M urea + 250 mM imidazole

(Figure IV-3B). Protein refolding was initiated with step-wise buffer exchange in the high-speed centrifuge using VirF storage buffer and decreasing concentrations of urea. Typically, the protein began to precipitate in the filtration unit at concentrations of urea less than 2 M.

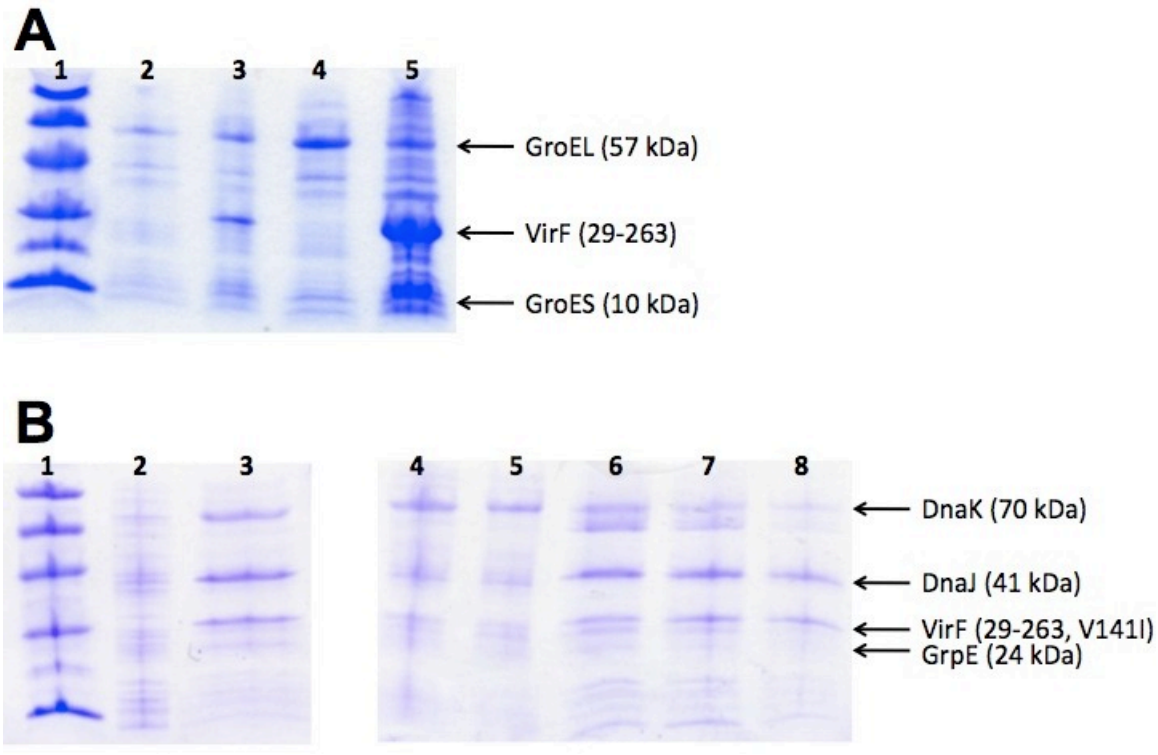


Figure IV-4: Co-expression of VirF(29-263) with Molecular Chaperones. A) Expression of wildtype VirF(29-263) with GroEL/ES complex at 37°C analyzed by SDS-PAGE and Coomassie Blue staining. Lane 1, low molecular weight standard; lane 2, uninduced control; lane 3, total protein (4h); lane 4, soluble fraction; lane 5, insoluble fraction. B) Expression of VirF(29-263, V141I) with DnaK/J and GrpE at 16°C and purification by Ni-NTA affinity chromatography analyzed by SDS- PAGE and Coomassie Blue staining. Lane 1, low molecular weight standard; lane 2, uninduced sample; lane 3, total protein (24 h); lane 4, flow-through; lanes 5-8, elute fractions.

As the solubility issue may be linked to protein mis-folding or aggregation, the VirF(29-263) protein was expressed in the presence of various molecular chaperones from Takara. Expression of the chaperone complex GroEL-GroES

(from pGro7), which functions at a later stage of protein folding, was first monitored with the truncated VirF protein. Following induction, the VirF(29-263) protein was still observed in the insoluble fraction, and the major protein in the soluble fraction was the molecular chaperone (Figure IV-4A). Co-expression of the DnaK-DnaJ complex with GrpE (from pKJE7) was also monitored, all of which are believed to function very early in protein folding. Although a small amount of soluble VirF(29-263) was isolated following Ni-NTA affinity chromatography, the protein could not be separated from the molecular chaperones, which co-elute with the target protein (Figure IV-4B).

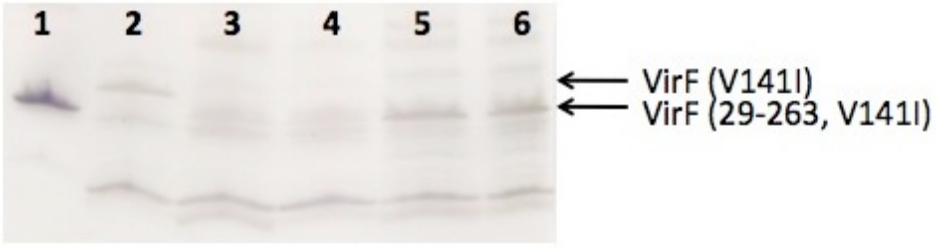


Figure IV-5: Expression of VirF(V141I) and VirF(29-263, V141I) with Molecular Chaperones. Co-expression of the (V141I) mutant full-length VirF and truncated VirF with DnaK/J and GrpE was monitored by anti-His Western blot. The soluble fraction from each was purified by Ni-NTA resin, and the elute fractions were analyzed by SDS-PAGE. Wildtype VirF(29-263) was run as a control. Lane 1, low molecular weight standard; lane 2, VirF(V141I); lanes 3-4, VirF(29-263); lanes 5-6, VirF(29-263, V141I).

Expression of the VirF(29-263, V141I) protein fragment and full-length VirF(V141I) were also monitored in the presence of the pKJE7 chaperones (DnaK, DnaJ, and GrpE). Both proteins (with 10x His-tag) were purified by Ni-NTA affinity chromatography and visualized by anti-His Western blot (Figure IV-5). There was a slight increase in soluble VirF(V141I) concentration, but a more pronounced effect was seen with soluble VirF(29-263, V141I). Though the

concentration of soluble protein improved in the mutant, isolation of purified protein was not achieved.

Purification of soluble VirF(154-263) was more successful (Figure IV-6A). Though the dimerization domain has been deleted, a gel-mobility shift assay was performed with the VirF(154-263) protein and a series of overlapping double-stranded DNA oligomers covering the key binding region in the *virB* promoter (Figure IV-6B) [11]. Purified dsDNA fragments of the *virB* promoter were incubated with varying concentrations of the VirF C-terminal DNA-bindingdomain, and the formation of a protein-DNA complex was monitored by gel-mobility shift assay and ethidium bromide staining (Figure IV-6C). At the highest protein concentration tested, the most pronounced VirF(154-263)-DNA complexes were formed with fragments corresponding to the -130 to -61, -80 to -11, and -70 to -31 regions of the *virB* promoter.

In vitro Translation Trials with VirF

Expression of VirF was monitored *in vitro* using a variety of plasmid constructs and purified *virF* mRNA. The commercially-available and lab-prepared extracts used provide coupled transcription/translation for genes under the control of the T7 promoter. Though the kits recommended using supercoiled DNA as a starting material, experiments were also performed with mRNA generated by *in vitro* transcription in an attempt to increase VirF yield. The VirF expression constructs tested were altered to include fusion tags amenable for Western blot analysis (*i.e.,* V5 epitope and His-tag). Experiments were performed according to vendor protocol, but the production of VirF was not

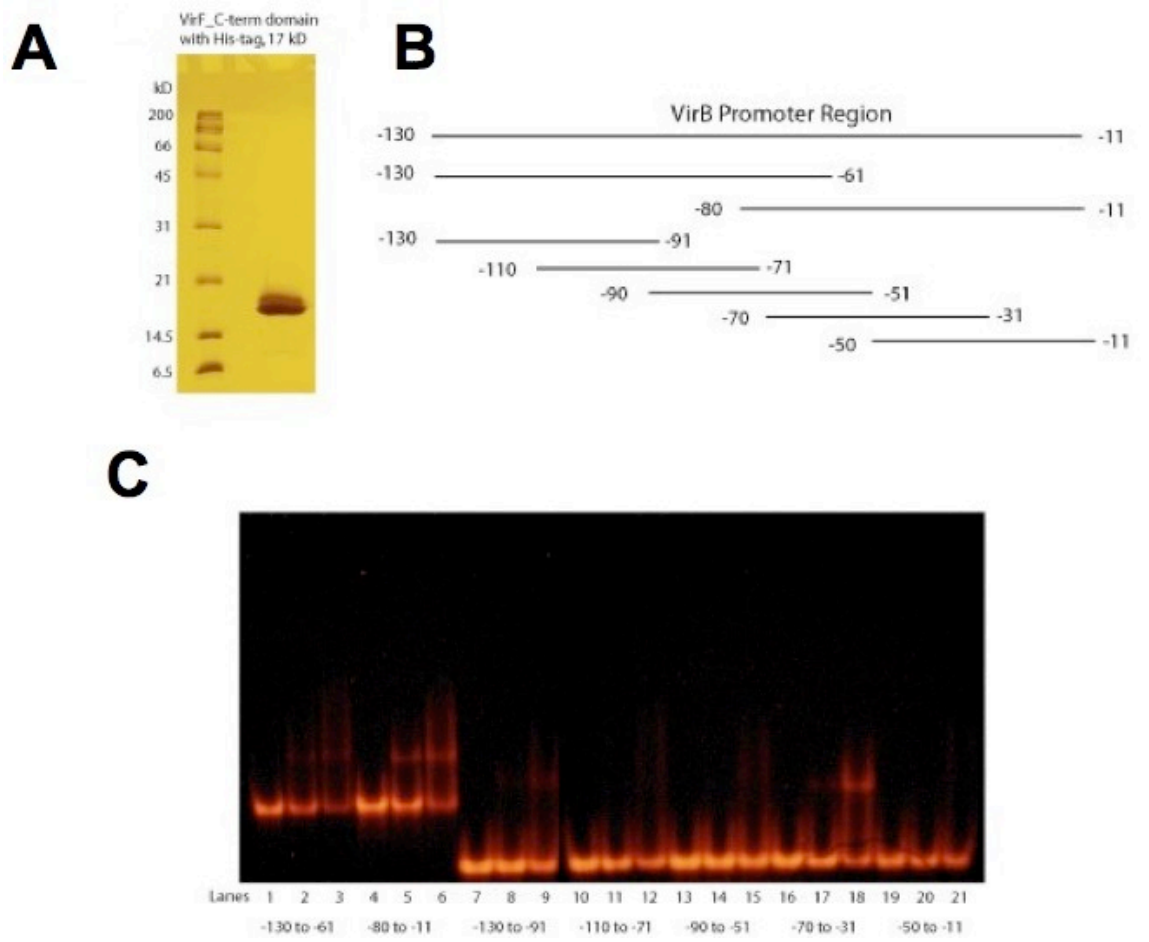


Figure IV-6: Purification and Analysis of VirF(154-263). A) Purified VirF(154-263) analyzed by SDS-PAGE and anti-His stain. B) Synthetic dsDNA fragments corresponding to regions the *virB* promoter from *Shigella*. C) *In vitro* gel-mobility shift assay with purified VirF(154-263) and dsDNA fragments analyzed by agarose gel and ethidium bromide staining.

detected by standard immunoblotting. Expression of the control proteins CAT and CALML-3 was only approximately 2-3 fold higher in comparison to background protein levels visualized by Coomassie Blue staining (a representative gel is shown in Figure IV-7).

For increased sensitivity, protein synthesis was measured by incorporation of L-[^{35}S]-methionine. In addition, a cell-free S12 extract was prepared using BL21(DE3) cells following induction of T7 RNA polymerase previously reported to

improve expression *in vitro* [18, 19]. As a quality control, the incorporation of L-[^{35}S]-methionine was measured in both negative (no DNA) and positive (chloramphenicol acetyltransferase, CAT) control reactions. Briefly, reactions were incubated for 4 hours at 37°C. Aliquots (10 µL) were removed and spotted on a glass-fiber filter following precipitation with NaOH (an untreated control was used to determine the total radioactivity in the reaction), and the observed radioactivity (CPM) was compared to the values predicted for the Ambion extract (ActivePro^{TM}) (Table IV-2). Though the signal to background ratio in the prepared extract is less than that observed in the commercial extract (~5-fold increase over the no DNA control), the sensitivity is improved in comparison with the immunoblotting detection method.

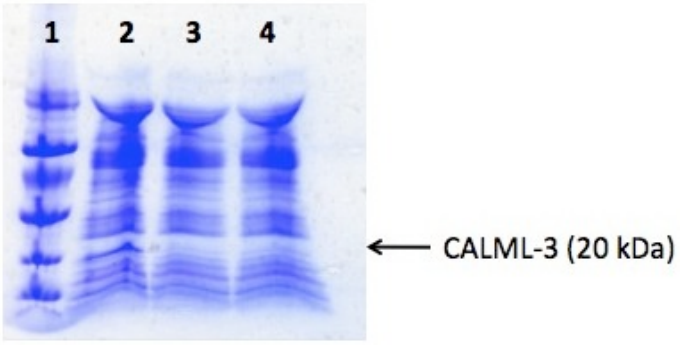


Figure IV-7: *In vitro* Translation of VirF Derivatives Analyzed by SDS-PAGE and Coomassie Blue Staining. Lane 1, low molecular weight standard; lane 2, CALML-3 control; lane 3, VirF+V5 at 30°C; lane 3, VirF+RBS+V5 at 37°C.

Expression from a variety of *virf*-containing plasmids was monitored in the presence and absence of *virB* DNA. With high background radioactivity in the *in vitro* translation reactions, only two of the samples tested demonstrated L-[^{35}S]-methionine incorporation over the no DNA control, VirF(154-263) and MalE-VirF. The CPM incorporation determined after 4 h for each VirF construct was low in

comparison to the CAT positive control (about 3-fold and 6-fold, respectively) (Figure IV-8). Precipitate routinely developed in the *in vitro* translation reactions and produced inconsistent data between replicates.

Table IV-2. Predicted and Experimentally Determined Values for Incorporation of L-[^{35}S]-methionine with *in vitro* Translation Assays. Aliquots were tested after 4 hours.

Measured Aliquot	Ambion Prediction (CPM)	Experimental Value (CPM)
Total radioactivity	3.9×10^6	3.3×10^6
No DNA control	5×10^3	4.1×10^3
CAT Positive Control	7.2×10^5	1.9×10^4

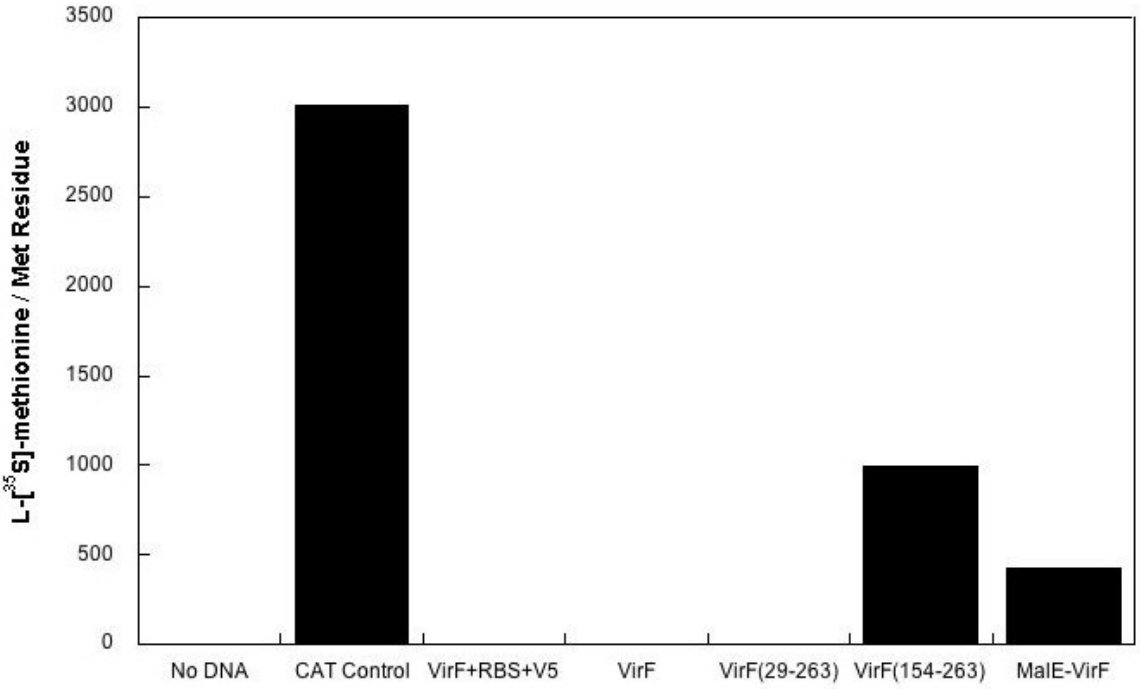


Figure IV-8: *In vitro* Translation of VirF Derivatives Monitored by L-[^{35}S]-methionine Incorporation. Chloramphenicol acetyltransferase (CAT) was monitored as a positive control and the no DNA sample as the negative control. Total protein samples were measured after a 4 h incubation period. The total radioactivity observed in each sample was normalized to the background radioactivity in the no DNA control (4100 CPM) and the number of methionine residues in each protein: CAT – 5, VirF(154-263) – 2, MalE-VirF – 11.

Discussion

VirF is a critical virulence factor in *Shigella spp.* As a member of the AraC family of transcriptional activators, VirF shares a conserved DNA-binding domain of 99 amino acids that are predicted to adopt two helix-turn-helix motifs (HTH) [12]. Some evidence suggests that the significant sequence homology in the second HTH indicates the importance of this domain in a conserved function of this diverse class of activators (*i.e.,* recruitment of RNA polymerase) [21]. In addition, VirF has an N-terminal domain, connected to the DNA-binding domain by a short linker region, consisting of more than 100 amino acids responsible for homodimerization [12]. Protein insolubility when expressed in *E. coli* has made structural analysis of VirF difficult [11, 12]. Several key insights into VirF structure and function have come from analyses of a MalE-VirF fusion protein and point mutants that resulted in expression of soluble protein. As always, the question remains as to whether or not the biologically relevant structure and function is maintained in these mutant forms.

With a wide variety of molecular biology tools available to manipulate protein expression, several fusion tags, inducible promoters, growth temperatures, and VirF protein fragments were considered in an attempt to optimize VirF expression. Uninduced maintenance of *virf* plasmids in *E. coli* is less problematic, as normal cell growth is observed and the plasmid can be obtained in standard miniprep procedures (See Chapter III). In the presence of inducer; however, VirF expressed from either the T7 or *lac* promoter or the tightly regulated ara_{BAD} promoter resulted in a mucoidy phenotype. It is unclear what

incites the observed mucoid production. One possibility is that VirF aggregation and/or improper folding triggers a stress response yielding the mucoidy phenotype. Protein misfolding can lead to formation of inclusion bodies, which disrupt the cytosolic environment and in some cases lead to cell death [22-24]. In some pathogenic species, mucoidy phenotype and biofilm formation is related to pathogenesis (*i.e., Pseudomonas*, *S. aureus*) [25, 26]. It is interesting to speculate that VirF could be activating a specific pathway linked to *E. coli* cell wall stability, although the successful expression of the active malE-VirF fusion suggests otherwise Unfortunately, the mucoidy phenotype prohibited purification and further characterization of full-length VirF.

Fusion tags have been shown to increase the solubility of target proteins. Though there are a number of common protein fusions, one that has been linked to protein folding is the *pelB* leader sequence. This peptide directs the target protein to the periplasm, where the signal sequence is cleaved by an endogenous peptidase, and the target protein is in some cases secreted from the periplasm into the culture medium [20]. There are many potential benefits with this tag. One important factor is the endogenous mechanism for cleavage of the protein fusion tag. Crystallography and other structural characterization techniques are performed following cleavage of all solubility or affinity tags, which in some cases can prove problematic. In addition, direction of folding to the periplasm eliminates the risk of intracellular toxic effects and potential proteolytic cleavage of the target protein. Other affinity tags were selected based on the ease of purification or detection in standard immunoblotting procedures. The

histidine tag (His-tag) is widely used in molecular biology for purification of target proteins by nickel affinity chromatography. However, in light of observed differential rates of protein expression based on queuine-cognate codon usage (see Chapter II), we also chose to express VirF as a fusion with the V5 epitope (GKPIPNPLLGLDST), which contains only two Q-cognate codons. Although there was visual improvement in cell phenotype and growth, no appreciable concentration of soluble full-length VirF protein was detected in the presence of any of these fusion tags.

Based on increased solubility with other monomeric members of the AraC family, we hypothesized that expression of truncated forms of VirF may be more successful *in vivo* [14]. In collaboration with Prof. Oleg Tsodikov and Dr. Tappan Biswas, plasmid constructs were prepared for expression of VirF with a deletion of 28 amino acids from the N-terminal domain, VirF(29-263), and the conserved C-terminal domain involved in DNA-binding, VirF(154-263). Overexpression of VirF(29-263) in *E. coli* produced a high concentration of insoluble protein (Figure IV-3). Following many attempts to improve solubility (*i.e.,* expression with molecular chaperones, denaturation/refolding procedures), purification of the VirF fragment proved difficult due to the very dilute nature of the soluble fraction. The difference in protein expression between full-length VirF and the VirF(29-263) fragment provide potential clues to the amino acids linked to insolubility. The abundance of insoluble VirF(29-263) protein detected and normal cell morphology indicates that one or more amino acids in this region of the N-terminal domain (residues 1-28) partially contribute to the toxic (e.g., mucoidy)

effects of VirF expression in *E. coli*. Though the N-terminal domain is largely uncharacterized, Porter and Dorman have demonstrated that Y23 is highly conserved among other AraC-type activators (*i.e.,* Rns, AggR), and the VirF(Y23S) mutant was inactive in a transcriptional activation assay [12]. It is interesting to note that more soluble protein was recovered when the VirF(29-263, V141I) protein was co-expressed with molecular chaperones than the wildtype protein (Figure IV-5). The mRNA contains a conservative purine to purine point mutation ($G_{421}A$), with a very conservative corresponding amino acid substitution, which eliminates recognition of the *virF* mRNA by TGT (at least *in vitro*) (see Chapter III for details). The observation that the soluble concentration of this conservative mutant is higher than the wildtype form may speak to the biological relevance of the guanine 421 modification site. Further studies to characterize this mutant are pending.

A dramatic improvement in solubility was achieved with truncation of the entire N-terminal domain of VirF, expressing just the DNA-binding domain of VirF(154-263). Though the identity of DNA-binding sites in the *virB* promoter was characterized previously with the MalE-VirF fusion *in vitro*, similar fragments were reanalyzed with the untagged C-terminal domain [11]. In the original experiment, DNA foot-printing analysis and transcriptional activation was monitored with various deletion mutants of the *virB* promoter (-120 to +10 region). With the truncated VirF(154-263) protein, two 70 bp fragments and five 40 bp fragments of the -130 to -11 region were analyzed to monitor overlapping regions of the *virB* promoter region (Biswas and Tsodikov, unpublished).

Incubation with varying concentrations of the DNA-binding domain of VirF and the linear DNA fragments demonstrated two binding sites: one in or upstream of the -35 region (between -70 and -31, relative to the transcription start site) and the other is further upstream, between -130 and -91 (Figure IV-7C). The observation that VirF(154-263) does not bind to the -90 to -51 and the -50 to -11 fragments is suggestive that the binding site most likely includes -50. Likewise, for the -130 to -91 binding site, we can deduce that it must include or lie upstream of -110 owing to the fact that C-VirF does not bind to the -110 to -71 fragment. These observations are in line with the report by Tobe *et al.*, where DNA fragments containing nucleotides -105 to -17 were required for full transcriptional activation by MalE-VirF. One interesting result of these studies is the observation that VirF (at least this truncated form) can bind to linear dsDNA, in apparent conflict with the demonstrated requirement of supercoiled DNA for transcription activation. However, it is possible that although VirF(154-263) can bind to linear DNA , supercoiled DNA is required to adopt an active conformation for dimerization and recruitment of RNA polymerase.

Because formation of inclusion bodies is not an issue *in vitro*, we attempted expression of several VirF constructs in cell-free extracts. Originally, we hoped to test the translation of VirF *in vitro* from unmodified and queuine-modified *virF* mRNA. We hoped to monitor any changes in protein translation between the two states of the *virF* mRNA (modified vs. unmodified), and thereby directly assess the biological relevance of the TGT modification in *virF*. Unfortunately, we were unable to generate modified *virF* mRNA due to a lack of

synthetic Q nucleoside at that time (see Chapter III for details). VirF expression was initially monitored from both DNA and mRNA starting materials in optimized cell-free extracts for synthesis of T7-driven constructs. Though the concentration of the target protein should be high in comparison to the total cellular protein mixture, the limit of detection is low with SDS-PAGE and Coomassie Blue staining. Constructs were tagged with either the V5 epitope or His-tag in an attempt to improve solubility and provide a more sensitive method of detection by Western blot. The appropriate sensitivity was ultimately achieved with detection of protein synthesis by L-[^{35}S]-methionine incorporation. Expression of the control protein, CAT, was readily identified by *in vitro* translation (~3,000 counts above background); however, syntheses of the various forms of VirF tested were quite low. The two constructs that demonstrated incorporation of labeled methionine above background were the two proteins that were expressed in soluble form in *E. coli*: VirF(154-263) and the MalE-VirF fusion (Figure IV-8). These results suggest that there is an inherently insoluble or unstable nature in VirF, and there is strong evidence to suggest this characteristic is linked to the N-terminal domain.

Ultimately, the MalE-VirF fusion vector was constructed according to a previously described protocol [11]. Though isolation of native VirF from the maltose-binding protein fusion was unsuccessful, Tobe *et al*. reported transcriptional activation by the fusion protein *in vitro*. Structural analysis of the VirF protein was not possible, but we were still interested in a functional characterization of the transcriptional activator. The MalE-VirF fusion could be

used to screen for potential VirF inhibitors *in vitro* as a secondary confirmation protocol for compounds identified *in vivo*. For a complete list of plasmids constructed for the purpose of VirF expression, see Table IV-A1 in Appendix IV-1.

Conclusions

Many attempts to express and purify VirF in *E. coli* and *in vitro* were unsuccessful. Several common molecular biology techniques were employed, including expression of fusion tagged proteins, co-expression with chaperones, various culture media, and lowered growth temperatures. In most cases where expression of the full-length protein was monitored (excluding pBADvirF+His), cell growth was challenged, and a mucoidy phenotype was observed. Ultimately, our observations are in line with those of other researchers who have tried to purify recombinant VirF. In collaboration with Prof. Oleg Tsodikov and Dr. Tappan Biswas, two truncated fragments of the VirF protein were prepared. The DNA-binding domain of VirF, VirF(154-263), was expressed in soluble form and DNA binding was characterized. Though structural characterization of native VirF has failed, we turn our attention to a functional analysis of VirF to better understand this critical virulence factor through small molecule inhibition.

Notes to Chapter IV

I gratefully acknowledge Prof. Oleg Tsodikov and Dr. Tappan Biswas for their work in cloning and expressing the VirF fragments, in addition to characterization of VirF(154-263).

Abbreviations used: MBP, maltose-binding protein; Ni-NTA, nickel-nitrilotriacetic acid; CIP, calf-intestinal alkaline phosphatase; IPTG, isopropyl β-D-thiogalactoside dNTP, deoxyribonucleotide triphosphate; HRP, horseradish peroxidase; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; Amp, ampicillin; CN/DAB, 4-chloro-1-naphthol/3,3-diaminobenzidine; PCR, polymerase chain reaction; RBS, ribosome-binding site; TOPO, topoisomerase; Kan, kanamycin; His, histidine; dsDNA, double-stranded DNA; bp, base pair; pps, PreCission protease site; EDTA, ethylenediaminetetraacetic acid; PBST, phosphate-buffered saline with Tween; DTT, dithiothreitol; PMSF, phenylmethylsulfonyl fluoride; Ni-IMAC, nickel-immobilized metal affinity chromatography; CAT, chloramphenicol acetyltransferase; CALML-3, calmodulin-like protein 3; ATP, adenosine triphosphate; FBP, fructose-1,6-bisphosphate; HEPES-KOH, 4-(2-hydroxyethyl-1-piperazineethanesulfonic acid); NTP, ribonucleotide triphosphate; PEG, polyethylene glycol; TCA, trichloroacetic acid.

Appendix IV-1

Table IV-A1. Summary of VirF Expression Plasmid Construction.

Construct Name	Solubility Tag?	Primary Use	Expression in Cells?
pTZ19R-virF (pTZvirF)	--	*In vitro* transcription	XL-2 Blue, TG2 Mucoidy
pTZvirF+RBS	--	*In vitro* translation	XL-2 Blue, TG2 Mucoidy
pTZvirF+RBS+ V5 (pTZvirF+RV)	V5	*In vitro* translation	BL21(DE3) Normal cultures, No expression
pTZvirF($G_{421}A$)	--	Test predicted modification site	TG2 (ND)
pET28a-virF	6x-His	*Lac* controlled system/ His-tagged	TG2 Mucoidy
pET20b-virF	PelB	Periplasmic expression	BL21(DE3), K12(Δ*tgt*) Normal cultures, No expression
pET20b-virF(29-263)	PelB	Periplasmic expression, begins at Met 29	BL21(DE3), K12(Δ*tgt*) Normal cultures, No expression
pET19b-virF	10x-His	Full-length His-tagged	BL21(DE3) (Biswas) Normal cultures, No expression
pET19b-virF (29-263)	10x-His	Cytoplasmic expression, begins at Met 29	BL21(DE3) (Biswas) Normal cultures, Insoluble protein
pET19b-virF (154-263)	10x-His	Expression of DNA-binding domain	BL21(DE3) (Biswas) Normal cultures, Soluble protein
pBDG302	--	PCR *virf* gene (tac promoter)	*S. flexneri* 2a (Björk), TG2 (ND)
pLacT7-virF	--	More tightly regulated promoter	XL-2 Blue Mucoidy, burst colonies (rich and minimal media)
pBAD/D-TOPO-virF (pBADvirF)	6x-His	ara_{BAD} promoter, His-tagged	TG2, BL21(DE3), LMG194, K12(Δ*tgt*) Mucoidy (rich / minimal media), Small bkgd. (Mass Spec.)
pMAL-c2x-virF (pMALvirF)	MalE (MBP)	Maltose-binding protein fusion	BL21 Normal cultures Soluble Protein

1. Ming-Shan, Q., H. Yoshikura, and H. Watanabe, *Virulence phenotypes of Shigella flexneri 2a avirulent mutant 24570 can be complemented by the plasmid-coded positive regulator virF gene.* FEMS Microbiology Letters, 1992. 92: p. 217-222.

2. Sasakawa, C., et al., *Molecular alteration of the 140-megadalton plasmid associated with loss of virulence and congo red binding-activity in Shigella flexneri.* Infection and Immunity, 1986. 51(2): p. 470-475.

3. Adler, B., et al., *A dual transcriptional activation system for the 230 kb plasmid genes coding for virulence-associated antigens of Shigella flexneri.* Molecular Microbiology, 1989. 3(5): p. 627-35.

4. Hale, T.L., *Genetic basis of virulence in Shigella species.* Microbiological Reviews, 1991. 55(2): p. 206-24.

5. Tobe, T., et al., *Temperature-regulated expression of invasion genes in Shigella flexneri is controlled through the transcriptional activation of the virB gene on the large plasmid.* Molecular Microbiology, 1991. 5(4): p. 887-893.

6. Porter, M.E. and C.J. Dorman, *A role for H-NS in the thermo-osmotic regulation of virulence gene expression in Shigella flexneri.* Journal of Bacteriology, 1994. 176(13): p. 4187-4191.

7. Tobe, T., M. Yoshikawa, and C. Sasakawa, *Thermoregulation of virB transcription in Shigella flexneri by sensing of changes in local DNA superhelicity.* Journal of Bacteriology, 1995. 177(4): p. 1094-7.

8. Falconi, M., et al., *Thermoregulation of Shigella and Escherichia coli EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of virF promoter to transcriptional repressor H-NS.* EMBO Journal, 1998. 17(23): p. 7033-43.

9. Prosseda, G., et al., *A role for H-NS in the regulation of the virF gene of Shigella and enteroinvasive Escherichia coli.* Research in Microbiology, 1998. 149(1): p. 15-25.

10. Prosseda, G., et al., *The virF promoter in Shigella: more than just a curved DNA stretch.* Molecular Microbiology, 2004. 51(2): p. 523-37.

11. Tobe, T., et al., *Transcriptional control of the invasion regulatory gene virB of Shigella flexneri: activation by virF and repression by H-NS.* Journal of Bacteriology, 1993. 175(19): p. 6142-9.

12. Porter, M.E. and C.J. Dorman, *In vivo DNA-binding and oligomerization properties of the Shigella flexneri AraC-like transcriptional regulator VirF as identified by random and site-specific mutagenesis.* Journal of Bacteriology, 2002. 184(2): p. 531-9.

13. Soisson, S.M., et al., *Structural basis for ligand-regulated oligomerization of AraC.* Science, 1997. 276(5311): p. 421-425.

14. Kwon, H.J., et al., *Crystal structure of the Escherichia coli Rob transcription factor in complex with DNA.* Nature Structural Biology, 2000. 7(5): p. 424-30.

15. Chong, S. and G.A. Garcia, *A Versatile and General Prokaryotic Expression Vector, pLACT7.* BioTechniques, 1994. 17(4): p. 686-691.

16. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expression and Purification, 2005. 41: p. 207-234.

17. Hurt, J.K., S. Olgen, and G.A. Garcia, *Site-specific modification of Shigella flexneri virF mRNA by tRNA-guanine transglycosylase in vitro.* Nucleic Acids Research, 2007. 35(14): p. 4905-13.

18. Kim, T.W., et al., *An economical and highly productive cell-free protein synthesis system utilizing fructose-1,6-bisphosphate as an energy source.* Journal of Biotechnology, 2007. 130(4): p. 389-393.

19. Kim, T.W., et al., *Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system.* Journal of Biotechnology, 2006. 126(4): p. 554-561.

20. Oelschlaeger, P., et al., *Identification of factors impeding the production of a single-chain antibody fragment in Escherichia coli by comparing in vivo and in vitro expression.* Applied Microbiology and Biotechnology, 2003. 61(2): p. 123-132.

21. Gallegos, M.T., et al., *Arac/XylS family of transcriptional regulators.* Microbiology & Molecular Biology Reviews, 1997. 61(4): p. 393-410.

22. de Groot, N.S., et al., *Studies on bacterial inclusion bodies.* Future Microbiology, 2008. 3(4): p. 423-435.

23. de Groot, N.S. and S. Ventura, *Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates.* Journal of Biotechnology, 2006. 125(1): p. 110-113.

24. Gonzalez-Montalban, N., et al., *Bacterial inclusion bodies are cytotoxic in vivo in absence of functional chaperones DnaK or GroEL.* Journal of Biotechnology, 2005. 118(4): p. 406-412.

25. Hall-Stoodley, L., J.W. Costerton, and P. Stoodley, *Bacterial biofilms: From the natural environment to infectious diseases.* Nature Reviews Microbiology, 2004. 2(2): p. 95-108.

26. Hall-Stoodley, L. and P. Stoodley, *Biofilm formation and dispersal and the transmission of human pathogens.* Trends in Microbiology, 2005. 13(1): p. 7-10.

CHAPTER V

Identification of Small Molecule Inhibitors of VirF: Development of a VirF – β-galactosidase Reporter Assay

Transcriptional activators of the AraC family are widely used as regulators of bacterial virulence (*i.e.,* Rns from *E. coli*, ExsA in *P. aeruginosa*, TcpN from *V. cholerae*, and VirF from *S. flexneri*). Commonly found in enteropathogens, AraC-type activators have also been identified in virulent bacteria that invade the respiratory and urinary tracts [1]. Separated into two classes, AraC family members that function as homodimers are regulated by either 1) small molecules or 2) physical stimuli [2, 3]. The third class of AraC transcriptional activators are monomeric and typically respond to cell stress [4]. Despite such differences in protein structure and regulation mechanisms, AraC family members share a high degree of sequence homology in approximately 100 amino acids that constitute the DNA-binding domain in the C-terminus.

Due to issues with protein solubility and expression of recombinant proteins (See Chapter IV for discussion of VirF), AraC family members have historically been difficult to study *in vitro* [5]. In the case of VirF from *Shigella flexeri*, molecular manipulations of both the *virf* gene and the downstream DNA target sequences have aided in our understanding of this critical virulence regulator. The binding sites within the *virB* promoter were identified for both VirF

and the negative regulator, H-NS, by DNA footprinting analysis [6]. In addition, key amino acids required for DNA-binding and protein oligomerization were identified by mutagenesis experiments [7]. However, it is still unclear how VirF activates transcription at the molecular level; understanding of this process is critical for development of potential small molecule inhibitors of VirF.

Much of the structural evidence to probe the DNA-binding properties of the AraC family comes from the monomeric transcriptional activators (including Rob and MarA) [8]. Recently, Paratek Pharmaceuticals reported identification of a class of hydroxybenzimidazoles that inhibited DNA-binding of MarA and other monomeric AraC-type activators *in vitro* (Figure V-1) [9]. A computational model of the DNA-binding domain of MarA was used to pre-screen small molecules *in silico,* and commercially available compounds were evaluated in a DNA-binding assay *in vitro.* With computational evidence and preliminary inhibition data, the authors took a directed synthetic approach to generate optimized inhibitors of this class of AraC proteins. Although DNA binding is critical for functionality of these activators, small molecules that inhibit the oligomerization of AraC-type activators could also be widely applicable to many members of this protein family.

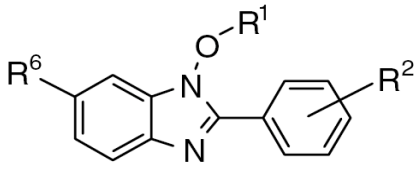


Figure V-1: Hydroxybenzimidazole Scaffold. Analogues were identified by Paratek Pharmaceuticals as inhibitors of MarA and other members of the AraC monomeric class of transcriptional activators.

Because expression of recombinant, dimeric AraC-type proteins has been problematic *in vitro*, we propose study of this class of transcriptional activators in a cell-based assay. Using a β-galactosidase (*i.e.,* LacZ) reporter assay in avirulent *Shigella flexneri* BS103, VirF-specific activation of the *virB* promoter (controlling transcription of the *lacZ* reporter gene) was monitored spectrophotometrically with the substrate CPRG (chlorophenol red β-D-galactopyranoside). Small molecules that demonstrated confirmed inhibition of the reporter assay were analyzed further in a dose response analysis.

Materials and Methods

Reagents

Unless otherwise specified, all reagents were from Sigma Aldrich. Gelase™ Agarose-Gel Digesting Preparation was from Epicentre. The QIAprep Spin Miniprep Kit was from Qiagen. Carbenicillin (disodium salt), Corning microtiter plates, bactotryptone and yeast extract were from Fisher Scientific. Electrocompetent NovaXG Zappers and Induction Control E BL21(DE3) were from Novagen. Chlorophenol red β-D-galactopyranoside (CPRG, monosodium salt) was from Roche. All restriction enzymes, Vent® DNA polymerase, Calf-intestinal alkaline phosphatase (CIP), and pACYC184 were from New England Biolabs. SeaPlaque agarose was from Cambrex. Isopropyl β-D-thiogalactoside (IPTG), 5-bromo-4-chloro-3-indolyl β-D-galactoside (X-gal), T4 DNA Ligase, pBAD202-D/TOPO and all synthetic oligonucleotides were from Invitrogen. The deoxynucleotide triphosphates (dNTPs, monosodium salts) were from Promega.

The electroporation 0.2 cm cuvettes were from BioRad. The permeable *E. coli* strain, EC2880, was a generous gift from Dr. Michael Hubband (Pfizer Scientific). The virulence plasmid-cured *S. flexneri* strain, BS103, was a generous gift from Prof. Anthony Maurelli (Uniformed Services University of the Health Sciences). The plasmid harboring *virB* was a generous gift from Prof. Chihiro Sasakawa (The University of Tokyo, Japan).

VirF-LacZ Reporter Plasmid Construction

Construction of Altered lacZα (ΔMCS)

The *lacZ α*-peptide (*i.e.,* β-galactosidase α-compliment) from the cloning vector pTZ18U^{Amp} was used as the template for subsequent plasmid construction. For cloning purposes, the multiple cloning site (MCS) were replaced by a 15 b.p. DNA cassette (see pTZ18U(ΔMCS) in Table V-I for oligo sequences). The complimentary oligos were designed with the appropriate free 3' and 5' ends as if following restriction digestion with *Eco*RI and *Hind*III. In a reaction volume of 33 µL, 0.75 nmol of each oligo was incubated at 37°C for 1h in 1x T4 DNA Ligase buffer (50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 1 mM ATP, 1 mM DTT, 5% PEG-8000). The dsDNA sample was precipitated overnight with 0.1 volume 3M NaOAc (pH 5.3) and 2.5 volumes 100% EtOH at -20°C. The vector pTZ18U was linearized via double restriction digestion with *Eco*RI and *Hind*III (20 U each, 20 µL volume) at 37°C for 1h. The vector was gel-purified from Seaplaque agarose gel with Gelase™ according to vendor protocol. In a total volume of 20 µL, 6 µg of the DNA cassette and 0.01 µg of digested pTZ18U were incubated overnight at 17°C with T4 DNA ligase (2 U). The ligated sample

(pTZ18U(ΔMCS), 10 µL) was transformed into 250 µL of chemically competent TG2 cells and plated on L-Carb plates (100 µg/mL carbenicillin). Individual colonies were isolated, and 3 mL 2xTY cultures (with 100 µg/mL carbenicillin) were inoculated at 37°C with shaking. Plasmid was isolated via miniprep, and the correct sequence was confirmed by DNA sequencing (University of Michigan DNA Sequencing Core Facilities). TG2 cells harboring pTZ18U(ΔMCS) exhibited β-galactosidase activity when grown at 37°C on L-Carb plates supplemented with X-gal and IPTG (40 µg and 3 µmol, respectively).

pACYC184 Reporter Dervatives

To generate the low-copy number counter screen, p*lac*-LacZ*α*, the *lac* operator and *lacZα* sequence were amplified by PCR from pTZ18U(ΔMCS). In a total volume of 54 µL, reactions containing 500 ng pTZ18U(ΔMCS), PCR primers (20 pmol each, see p*lac*-LacZ*α* in Table V-1), 1.85 mM $MgSO_4$, dNTP mix (0.24 mM each NTP) and Vent® DNA polymerase (2U) were incubated according to the following temperature sequence: 30 cycles – 94°C for 1 min., 45°C for 1 min., and 72°C for 2 min. The amplified *lac*-lacZα sequence and pACYC184 were treated in double restriction digestions with *Ava*I and *Bam*HI (40 U each enzyme, 20 µL reaction) at 37°C for 1 h. Both the plasmid and PCR product were purified, ligated, and transformed into TG2 cells as previously described. The resultant plasmid, p*lac*-LacZ*α*, was isolated via miniprep from 3 mL 2xTY cultures (with 30 µg/mL chloramphenicol), and samples were confirmed with DNA sequencing.

For the VirF-driven LacZ *α*-peptide reporter, p*vir*-LacZ*α*, the *virB* operator was amplified from pTB601 [6] and ligated with *lacZ(ΔMCS) α*-compliment in two

Table V-1: Oligonucleotide Sequences for Reporter Plasmid Construction. Key restriction enzyme sites or nucleotides of interest are underlined.

Sample Name	Primer Name	Primer Sequence (5' – 3')
pTZ18U (ΔMCS)	*Eco*RI ΔMCS FWD	AATTCGACTGCGCGCGAGTCA
	*Hind*III ΔMCS REV	AGCTTGACTCGCGCGCAGTCG
p*lac*-LacZα	LacZα *Ava*I FWD	CTGTCCCGGGGAGCGCCCAATACG
	LacZα *Bam*HI REV	GCTGGATCCCCATCACCCTAATC
virB Op	virB Op *Ava*I FWD	AGCTCCCGGGCTCACATCAGAGCTCC
	virB Op *Aat*II REV	AGCTGACGTCATCACACCCTGTTTATTC
lacZα	LacZα *Aat*II FWD	GAGACGTCATGACCATGATTAC
	LacZα *Bam*HI REV	*See sequence above
p*vir*-LacZα + RBS-6x His	Add RBS-His FWD	CAGACGTCAAGGAGATATACCATATG-CATCATCATCATCATCACGACGTCAC
	Add RBS-His REV	GTGACGTCGTGATGATGATGATGATG-CATATGGTATATCTCCTTGACGTCTG
LacZ(Δ*Aat*II) Quik-Change	LacZα *Aat*II QC FWD	GATGAGCGGCATTTTCCGTGACGTAT-CGTTGCTGCATAAACCGACTACAC
	LacZα *Aat*II QC REV	GTGTAGTCGGTTTATGCAGCAACGAT-ACGTCACGGAAAATGCCGCTCATC
pMALvirF-LacZα	virB Op *Pst*I FWD	GCTCTGCAGCTCACATCAGAGCTCC
	LacZα *Xba*I REV	CTTCTAGACTAATCAAGTTTTTTGG
pMALvirF -LacZ	LacZ *Aat*II FWD	GAGGACGTCATGGGCAGCAGCCATC
	LacZ *Xba*I REV	GAGTCTAGATTATTTTTGACACCAG
pBADvirF -LacZ	MalE-VirF-LacZ FWD	CACCATGAAAATCGAAGAAG
	MalE-VirF-LacZ REV	CAGTTTAAACCTCACATCAGAG

forms: 1) separated by only the *Aat*II restriction enzyme site or 2) with the *E. coli* ribosome-binding site (RBS) and 6x His-tag upstream of the *lacZ α* start codon. Briefly, in a volume of 53 µL, samples were amplified in triplicate containing 500 ng pTB601, PCR primers (20 pmol each, see *virB* Op in Table V-1), dNTP mix (0.24 mM each NTP), and Vent® DNA polymerase (2U) as previously described. The *lacZ(∆MCS)* fragment was amplified from pTZ18U(∆MCS) under the same conditions with the appropriate primers (see *lacZα* in Table V-1). All replicate reactions were combined and ethanol precipitated at -20°C overnight. Each DNA precipitate was resuspended in water (40 µL). Both PCR products (20 µL DNA, 40 µL total volume) were treated with *Aat*II (40 U) at 37°C for 1 h. The digested samples were gel-purified as previously described. The *virB* Op and *lacZ(∆MCS)* fragments were ligated overnight at 16°C (2:1 volume ratio, 20 µL total volume) with T4 DNA Ligase (2 U). To increase the concentration, the ligated product was amplified by PCR under the following conditions in duplicate: *virB Op*-lacZα (1 µL ligation sample), dNTP mix (0.25 mM each NTP), virB Op *Ava*I FWD primer (20 pmol), LacZα *Bam*HI REV primer (20 pmol), and Vent® DNA polymerase (2U) in a 52 µL volume as previously described. The reactions were combined and ethanol precipitated at -20°C overnight. The DNA pellet was resuspended in water (20 µL). The amplified *virB Op*-lacZα fragment was subcloned into pACYC184 following double restriction digestion with *Ava*I and *Bam*HI according to the protocol described for p*lac*-LacZα.

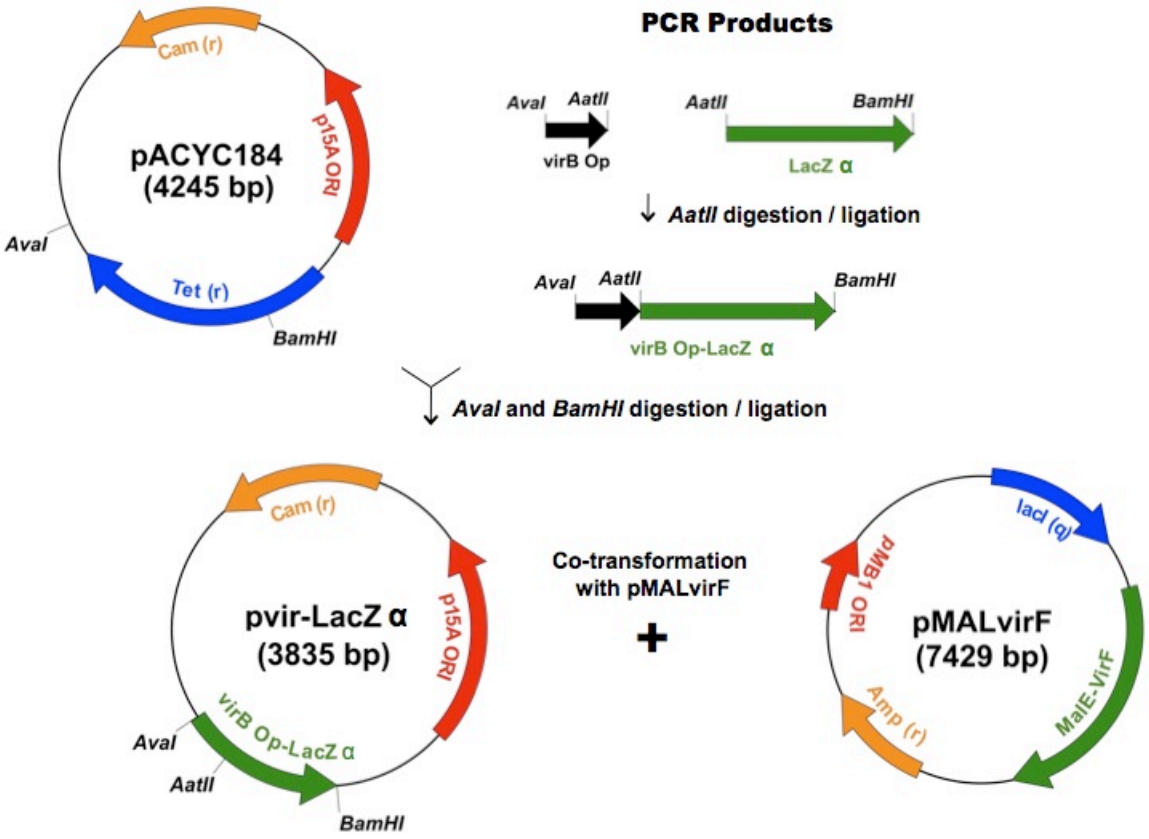


Figure V-2: Construction of the Dual-Plasmid Reporter with LacZ α-peptide.

Following construction of p*vir*-LacZα, a dsDNA cassette was engineered (see p*vir*- LacZα + RBS-6x His in Table V-1) to incorporate the *E. coli* ribosome-binding site concensus sequence (RBS, AGGAGA) and a 6x-Histidine tag for expression trials in *E. coli*. Designed with *Aat*II restriction sites on either end, the two complimentary oligos (0.1 µmol each) were annealed in water at 37°C for 1 h. The dsDNA insert and p*vir*-LacZα (20 µL DNA) were treated with *Aat*II (40 U, 26 µL total volume) at 37°C for 1 h. Alkaline phosphatase (CIP, 30 U) was added directly to the digested p*vir*-LacZα sample to prevent religation, and the sample was incubated at 37°C for an additional 1 h. Both digested fragments were gel-purified as previously described. In a total volume of 20 µL, the dsDNA insert and

digested p*vir*-LacZα (5:1 volume ratio) were incubated overnight at 17°C with T4 DNA ligase (2 U). The ligated sample (p*vir*-LacZα+RH, 10 µL) was transformed into 250 µL of chemically competent TG2 cells and plated on L-Carb plates (100 µg/mL carbenicillin). Individual colonies were isolated, and 3 mL 2xTY cultures (with 100 µg/mL carbenicillin) were inoculated at 37°C with shaking. Plasmid was isolated via miniprep (Qiagen).

Electrocompetent TG2 cells (50 µL) were transformed with p*vir*-lacZ*α* (or p*vir*-lacZ*α*+RH) and pMALvirF (see Chapter IV) simultaneously, according to the BioRad electrotransformation protocol. Cells were grown overnight on L-plates (with 100 µg/mL carbenicillin, 30 µg/mL chloramphenicol, 40 µg X-gal and 3 µmol IPTG) at 37°C. Plasmid DNA was isolated from the positive (blue) transformants. A representative scheme for this dual VirF reporter subcloning strategy is illustrated in Figure V-2.

High-copy Number Reporter Plasmids

A collection of single, high-copy number vectors was also constructed. The *virB Op-lacZα* fragment was initially amplified by PCR and subcloned into pMALvirF. Briefly, in a total volume of 50 µL, samples containing 500 ng p*vir*-LacZα, 2 mM $MgSO_4$, dNTP mix (2.5 mM each NTP), PCR primers (20 pmol each, see pMALvirF-LacZα in Table V-1), and Vent® DNA polymerase (4 U) were incubated according to the following temperature sequence: 30 cycles – 94°C for 30 sec., 50°C for 1 min., and 72°C for 2 min. The amplified *virB Op-lacZα* sequence and pMALvirF (15 µL DNA) were treated in double restriction digestions with *Pst*I and *Xba*I (40 U each enzyme, 30 µL reaction) at 37°C for 1

h. Both the plasmid and PCR product were purified, ligated (7:1 volume ratio, insert to plasmid), and transformed into TG2 cells as previously described. The resultant plasmid, pMALvirF-LacZα, was isolated via miniprep from 3 mL 2xTY cultures with carbenicillin, and samples were confirmed with DNA sequencing.

Full-length *lacZ* (3 kb) was amplified by PCR from the Novagen Induction Control E (pET28a derivative). An internal *Aat*II restriction site was altered with a silent mutation introduced by site-directed mutagenesis (according to Stratagene protocol). In a 30 µL total volume, samples containing 500 ng pET28a-lacZ, Quik-change PCR primers (10 pmol each, see LacZ(Δ*Aat*II) Quik-Change in Table V-1), dNTP mix (0.25 mM each NTP), 2 mM MgSO4, and Vent® DNA polymerase (2 U) were incubated according to the following temperature sequence: 15 cycles – 94°C for 1 min., 50°C for 1 min., and 72°C for 6.5 min. The amplified samples were treated with *Dpn*I (40 U) in 1x NEBuffer 4 (20 mM Tris-Acetate (pH 7.9), 50 mM K(CH_3COO), 10 mM Mg($CH_3COO)_2$, 1 mM DTT) at 37°C for 2 h to digest wildtype DNA. The digested sample (10 µL) was transformed into 250 µL chemically competent TG2 cells and grown overnight on L-Kan plates (50 µg/mL kanamycin). Plasmid DNA was isolated, and the sequence was confirmed as previously described.

The corresponding *lacZ* open reading frame was amplified in triplicate as follows (30 µL volume): 5 µg pET28a-LacZ(Δ*Aat*II), PCR primers (10 pmol each, see pMALvirF-LacZ in Table V-1), dNTP mix (0.25 mM each NTP), 2 mM MgSO4, and Vent® DNA polymerase (2 U) were incubated according to the following temperature sequence: 30 cycles – 94°C for 1 min., 55°C for 1 min.,

and 72°C for 2 min. The samples were combined and ethanol precipitated at -20°C overnight. The DNA pellet was resuspended in water (30 µL). The *lacZ* gene was subcloned into pMALvirF-LacZ*α* via a double restriction digestion. Briefly, samples containing 5 µL DNA (10 µL total volume) were treated with *Aat*II and *Xba*I (20 U each) at 37°C for 1.5 h. Both the plasmid and PCR product were gel-purified, ligated (5:1 volume ratio, insert to plasmid), and transformed into TG2 cells as previously described. The resultant plasmid, pMALvirF-LacZ, was isolated via miniprep from 3 mL 2xTY cultures with carbenicillin, and samples were confirmed with DNA sequencing. The entire *malE-virF+virB Op-lacZ* fragment (~6 kb) was amplified for insertion by TOPO cloning into the ara_{BAD}-controlled vector, pBAD202-D/TOPO, according to vendor protocol. Briefly, samples containing 500 ng pMALvirF-LacZ, 2 mM $MgSO_4$, dNTP mix (0.25 mM each NTP), PCR Primers (20 pmol each, see pBADvirF-LacZ in Table V-1), and Vent® DNA polymerase (2 U) in a total volume of 50 µL were incubated according to the following temperature sequence: 30 cycles – 94°C for 1 min., 50°C for 1 min., and 72°C for 2 min. TOPO cloning was conducted according to the electrocompetent *E. coli* transformation protocol (Invitrogen). The pBAD202-D/TOPO sample was incubated with 12 ng PCR product for 30 min. at room temperature, and the ligation product (pBADvirF-LacZ, 3 µL) was transformed into electrocompetent NovaXG Zapper™ cells according to vendor protocol. Cells were grown at 37°C on L-Kan plates, and plasmid was isolated as previously described.

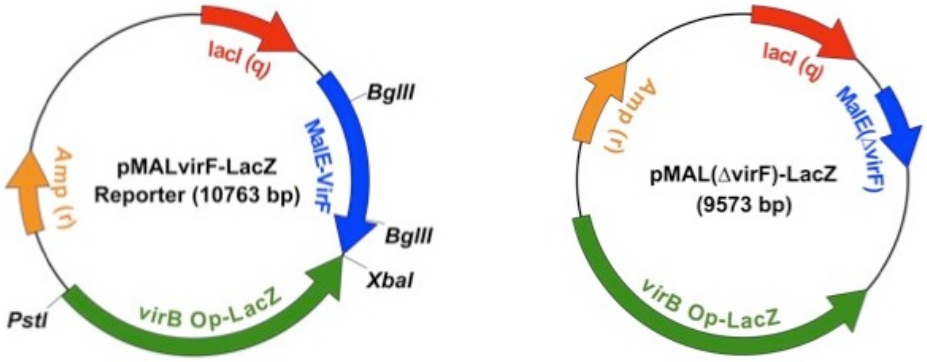


Figure V-3: Single Plasmid Reporter Constructs with LacZ.

A positive control vector was created for both pMALvirF-LacZ and pBADvirF-LacZ as a measure of β-galactosidase activity in the absence of VirF. Each reporter plasmid (10 µL, 20 µL total volume) was treated with the restriction endonuclease *Bgl*II (20 U) at 37°C for 90 min. The digested plasmids were gel-purified as previously described to remove the resultant 1200 nucleotide fragment of the *malE-virF* open reading frame. The purified samples were re-ligated at 17°C overnight with T4 DNA ligase (2 U). The ligation products (5 µL) were transformed into 250 µL chemically competent TG2 cells, and grown at 37°C on L-Carb (pMAL reporter) or L-Kan plates (pBAD reporter). Representative plasmid diagrams with key restriction enzyme sites for pMALvirF-LacZ and pMAL(ΔvirF)-LacZ are shown in Figure V-3.

β-galactosidase Assay Protocols

MUG Assay

This approach is a modified Miller method for cell-based assays in microtiter plate format described by F. Vidal-Aroca and coauthors [10, 11]. This method was analyzed with the pACYC184 derivatives p*lac*-LacZα and p*vir*-LacZα (in the presence and absence of pMALvirF). Cells were grown in batch culture

(rich 2xTY and minimal M9 media tested) containing 30 μg/mL chloramphenicol at 37°C with shaking to mid-log phase (OD_{600} ≅ 0.6-0.7). Varying concentrations of IPTG (0-5 mM) were added, and samples were incubated at 37°C with shaking for an additional 4 h. Cultures were assayed in 96-well microtiter plates (both black/clear bottom and black/opaque were tested) for β-galactosidase activity following 2 h and 4 h of induction with IPTG. Briefly, cells were diluted 5-fold in plates with Miller's Z buffer without β-mercaptoethanol (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$). The cell density was measured at A_{595} (black/clear plates only). The samples were incubated at room temperature for 15 min. following addition of the methylumbelliferyl β-D-galactopyranoside substrate (MUG, 25 μg in DMSO). The reactions (125 μL total volume) were quenched with 1M Na_2CO_3 (30 μL), and accumulation of the 4-MU reaction product was measured by fluorescence (λ_{ex} = 360 nm, λ_{em} = 460 nm) in a Molecular Dynamics microtiter plate reader. Calculation of Miller "MUG" Units was performed with the following equation (black/clear plates only): Activity = $F_{360/460}/(t \bullet A_{595})$.

CPRG Assay

The lab protocol for assaying β-galactosidase activity with the substrate CPRG (chlorophenol red β-D-galactopyranoside) was modified from a protocol developed by Thomas McQuade at Pfizer Inc. Briefly, this assay was performed with α-complementation vectors (p*lac*-LacZα, pTZ18U(ΔMCS), and pTZ18U) and reporters expressing full-length LacZ (pET28a-LacZ, pMALvirF-LacZ, and pBADvirF-LacZ) in both *E. coli* (TG2 and EC2880) and *S. flexneri* (BS103). A

wide variety of conditions were tested: the concentration of cells (OD_{600} = 1.5-0.03), inducer ([IPTG] = 2-0.25 mM, [L-arabinose] = 0.02-0.00002%), substrate ([CPRG] = 0.25-0.5 mg/mL), detergent ([Triton X-100] = 0.05-0.01%), and growth conditions (plated/batch culture, growth at 37°C vs. 30°C, etc.). The optimized conditions for each construct is outlined below.

For alpha-complementation with the altered LacZ α-peptide, *E. coli* TG2 cells harboring either p*lac*-LacZα or pTZ18U(ΔMCS) were grown in 2xTY batch culture (supplemented with 30 µg/mL chloramphenicol or 100 µg/mL carbenicillin, respectively) to saturation (OD_{600} ≅ 5). Samples were diluted in 384-well microtiter plates with 2xTY containing the appropriate antibiotic to the following final concentrations (50 µL total volume): OD_{600} = 0.3, 0.6, and 1.2. In addition to an uninduced control, each sample was plated with 2.5 mM IPTG at 37°C. β-galactosidase activity was measured with the addition of 0.5 mg/mL CPRG and 0.05% Triton X-100 (equal volume ratio, 50 µL total volume). The plates were incubated at room temperature for 45-60 min. All samples were assayed in replicate (n=6) after 4 h and 24 h post-induction. Accumulation of the CPR reaction product was detected by measuring absorbance (λ_{max} = 575 nm).

Assays performed with constituitive expression of the unaltered LacZ α-peptide (pTZ18U in TG2 cells) and full-length LacZ (pET28a-LacZ in TG2 and EC2880) were conducted as described previously with a few alterations based on observations from the previous assays. Cells were grown in batch 2xTY culture with appropriate antibiotics at 37°C to OD_{600} ≅ 1.5 and plated with low concentrations of IPTG (0, 0.25, 0.5 mM) for 24 h at 37°C. CPRG was held

constant at 0.5 mg/mL and the concentration of Triton X-100 was varied (0.01-0.05%). Plates with TG2 cells were incubated at room temperature for 45-60 min., and the EC2880 samples were incubated for 20 min. at room temperature prior to absorbance measurements.

The assay conditions for the VirF reporter plasmids (pMALvirF-LacZ and pBADvirF-LacZ) in EC2880 were as previously described with the following exceptions. Cells were grown in batch 2xTY culture with appropriate antibiotics at 37°C to OD_{600} ≅ 0.7-0.8 and plated in a 30 µL total volume with high concentrations of IPTG (2 mM) or L-arabinose (0.02%). The pMAL samples were grown at 37°C for 18 h, and the pBAD samples were incubated for 4 h. Cells were diluted to a total volume of 50 µL with 2xTY prior to addition of the substrate/detergent mixture. CPRG and Triton X-100 concentration were held constant at 0.25 mg/mL and 0.05%, respectively. The CPR reaction product was measured following a 10 min. incubation at room temperature. The same constructs (in addition to pMAL(ΔvirF)-LacZ) were assayed in avirulent *S.flexneri* BS103 using essentially the same protocol. For each construct the concentration of inducer was varied (for pMAL reporters, [IPTG] = 0-2 mM; for the pBAD reporter, [L-arabinose] = 0-0.02%), and the induction period was shortened to 3 h at 37°C. The CPRG/Triton X-100 mixture was added, and absorbance was measured following a 10 min. incubation at room temperature.

CPRG Assay Conducted at the CCG

The protocol for measuring β-galactosidase activity from BS103 harboring pMALvirF-LacZ (or the positive control for inhibition of β-galactosidase activity,

pMAL(ΔvirF)-LacZ) was optimized at the Center for Chemical Genomics (CCG, University of Michigan). Cells were grown to saturation in 2xTY at 37°C with shaking. The pMALvirF-LacZ samples were diluted in 384-well microtiter plates with 2xTY supplemented with carbenicillin to a final OD_{600} = 0.004 (30 µL total volume). Plates were incubated for 24 hours at 30°C. Cell density (OD_{600}) was monitored prior to addition of the CPRG/Triton X-100 mixture (0.5 mg/mL and 0.05%, respectively). The samples were incubated at room temperature for 10 min. before measuring CPR absorbance (A_{570}) in the PHERAStar plate reader with a narrow band pass filter. Initially, assay consistency was monitored in the absence of compound (VirF reporter tested in replicate, n=352). Subsequent trials with small molecule libraries (~42,000 compounds) were conducted with the positive and negative controls assayed in replicate (n=32), and diluted cells grown for 24 h at 30°C in the presence of test compounds (10 µM, n=1) added to 20 µL 2xTY by pintool. Potential hits were considered active by the following criteria: ≥ 55% inhibition by plate of the VirF-LacZ assay in comparison to controls, and inhibition of cell growth (OD_{600}) ≤ 20%. This compound set (1273 small molecules) was confirmed according to the above assay protocol (n=3). In addition, the test set was pre-incubated with purified β-galactosidase in Z buffer (25 nM, 30 µL total volume) for 1 h at 30°C. This preliminary *in vitro* counterscreen was conducted in triplicate with the same CPRG/Triton X-100 mixture. The absorbance measurements were made following a 2.5 min. incubation at room temperature.

Following data triage, compounds with reactive or toxic functionalities, solubility or promiscuity issues were discarded. Remaining test compounds (238 small molecules) were assayed in a dose response study (n=4) according to the previous protocol with the following exception: the concentration of small molecule was varied in 2-fold serial dilutions, ranging from 100-1.5 μM. The data were plotted as a function of the concentration of small molecule (μM) versus absorbance (570 nm). The data were fit by non-linear regression to the following equation for sigmoidal kinetics:

$$y = \text{lower}+((\text{upper}-\text{lower})/(1+10^{(([I]-M1)*M2))})$$

where lower is the lower limit of the assay, and upper is the upper limit of the assay, [I] is the inhibitor concentration, M1 is the IC_{50} and M2 is the Hill Slope. During the fitting process, the upper limit was manually entered (from the no inhibitor control) owing to the fact that the data generally do not go to low enough concentrations to sufficiently define the upper portion of the curve. The lower limit was also manually entered as defined by the average absorbance of the positive control for each individual plate.

Results

Construction of β-galactosidase Reporter Plasmids

Several low-copy number pACYC184 derivatives were constructed for the purpose of controlled expression of the *lacZ(ΔMCS)* α-peptide. The VirF-specific reporter plasmid, p*vir*-LacZ*α* was constructed for cotransformation with the VirF expression vector, pMALvirF. The counterscreen for this system (p*lac*-LacZα)

measures constituitive expression of LacZ*α* from the *lac* promoter. In parallel, another line of plasmids was constructed as a single expression/reporter system. The *virB Op-lacZα* fragment was subcloned directly into pMALvirF, and the *malE-virF+virB Op-lacZα* fragment was inserted into pBAD202-D/TOPO. A counterscreen was generated via deletion of a portion of the *malE-virF* open reading frame, eliminating expression of VirF from these samples. Through α-complementation or expression of the full-length reporter, β-galactosidase activity from all plasmids was assessed via cell growth on X-gal/IPTG supplemented agar plates.

β-galactosidase α-complementation Assay Trials

The β-galactosidase assay conditions were first optimized with the vector p*lac*-LacZα. Independent cultures were grown in 2xTY with chloramphenicol at 37°C with shaking. Following induction with IPTG, aliquots were removed at various time points, and β-galactosidase activity was measured with MUG in a modified Miller assay. The optimal assay conditions for enzymatic activity in p*lac*-LacZα were determined to be a 4-hour induction period with 2.5 mM IPTG (Figure V-4A). Due to significant background fluorescence in the uninduced samples, a series of negative control samples were tested (Figure V-4B). There was a slight increase in MUG hydrolysis in the absence of inducer and/or the transcription factor, VirF, in all of the reporter negative control samples with respect to the media control. Interestingly, the background fluorescence of the 2xTY media itself was quite high (approximately 1500 fluorescence units). β-galactosidase expression was also monitored from p*lac*-LacZα in minimal M9

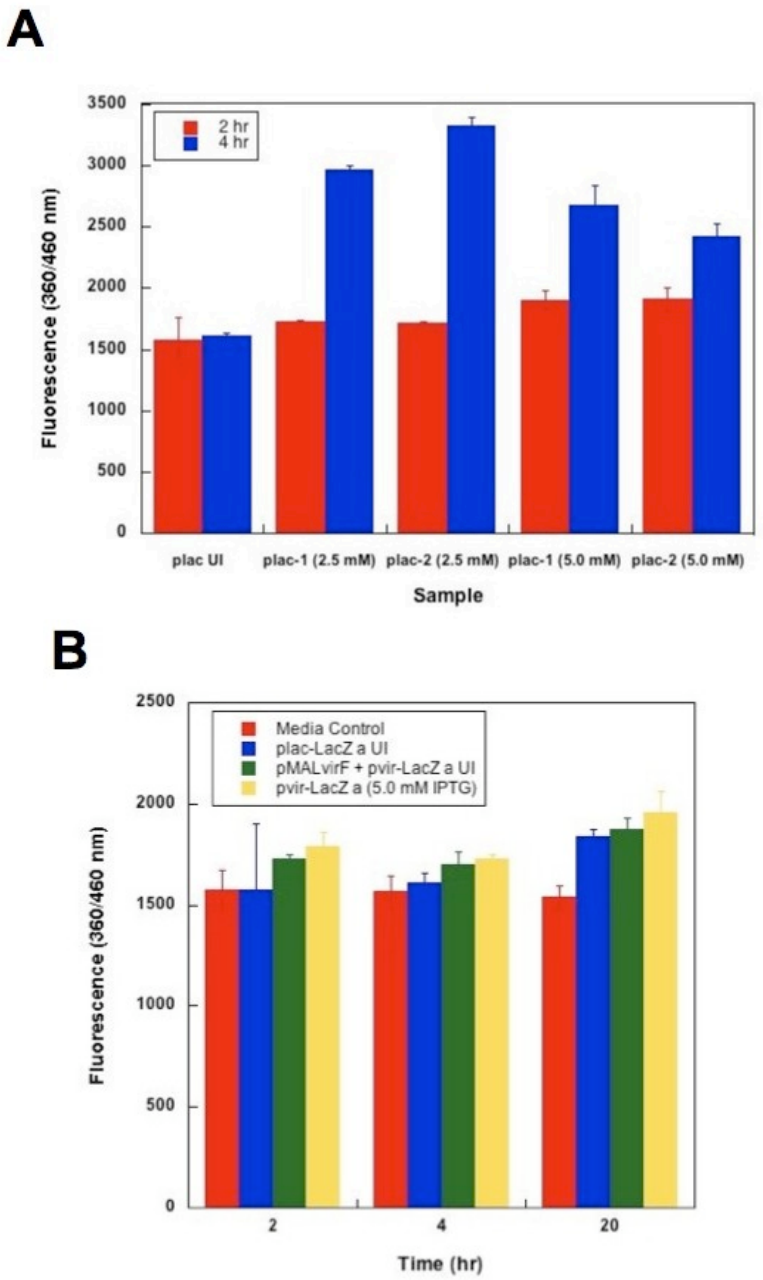


Figure V-4: β-galactosidase Assay Trials with MUG Substrate. A) MUG assay performed with plac-LacZ α (termed plac here) in *E. coli* TG2 in duplicate. Aliquots were removed from uninduced samples, 2 h (■) and 4 h (■) post-induction with IPTG (tested in triplicate). Detection of the 4-MU reaction product ($F_{360/460\ nm}$) was measured, and plotted as the average and standard deviation. B) Negative controls tested with the MUG substrate. The background fluorescence was measured (in triplicate) in 2xTY media only (■), uninduced plac-LacZ α (■), uninduced reporter pMALvirF + pvir-LacZ α (■), and pvir-LacZ α induced in the absence of pMALvirF (■).

media. Although the background fluorescence was reduced 3-fold, accumulation of the 4-methylumbelliferone (4-MU) reaction product also increased over time in the plasmid-free TG2 control samples (data not shown).

Due to issues with very high background signal with MUG, a more sensitive β-gal substrate was selected for further assay optimization. CPRG is spectrophotometrically active (A_{575}) and sensitive to low concentrations (nM range) of β-galactosidase. Conditions were first optimized in the altered *lacZ* α-peptide constructs, p*lac*-LacZα and pTZ18U(ΔMCS), to compare enzymatic activity in relationship to plasmid copy number. All samples were grown with shaking at 37°C in independent 2xTY cultures to late log-phase (OD_{600} ≅ 5). Varying the concentration of cells, samples were induced with 2.5 mM IPTG in 384-well clear microtiter plates and grown in the 37°C oven. Aliquots removed after a 4 hr induction period revealed little change in β-galactosidase activity via alpha-complimentation from p*lac*-LacZα between uninduced cells and those supplemented with IPTG (Figure V-5). As predicted, LacZ α-complementation with the high-copy number vector, pTZ18U(ΔMCS), was higher than that observed in p*lac*-LacZα. After 4 hours of growth at 37°C, pTZ18U(ΔMCS) demonstrated increased β-galactosidase activity in the absence of IPTG (2-4 fold higher than the induced cultures). Enzyme α-complementation was approximately 3-fold higher in pTZ18U(ΔMCS) in comparison to p*lac*-LacZα. For both constructs, overnight growth at 37°C in the presence of IPTG saturated the expression of LacZ.

Expression Trials with Full-length LacZ

Assays were also conducted to measure expression of LacZ via α-complementation from the wildtype vector pTZ18U in comparison to expression of full-length LacZ (Figure V-6). TG2 cells (*lacZ*Δ*M15*) were transformed with both pTZ18U and the pET28a derivative, pET-LacZ. Cells were plated (OD_{600} = 1.5) with varying concentrations of IPTG for 24 hours at 37°C in microtiter plates. The following day, the concentration of Triton X-100 detergent was also varied to

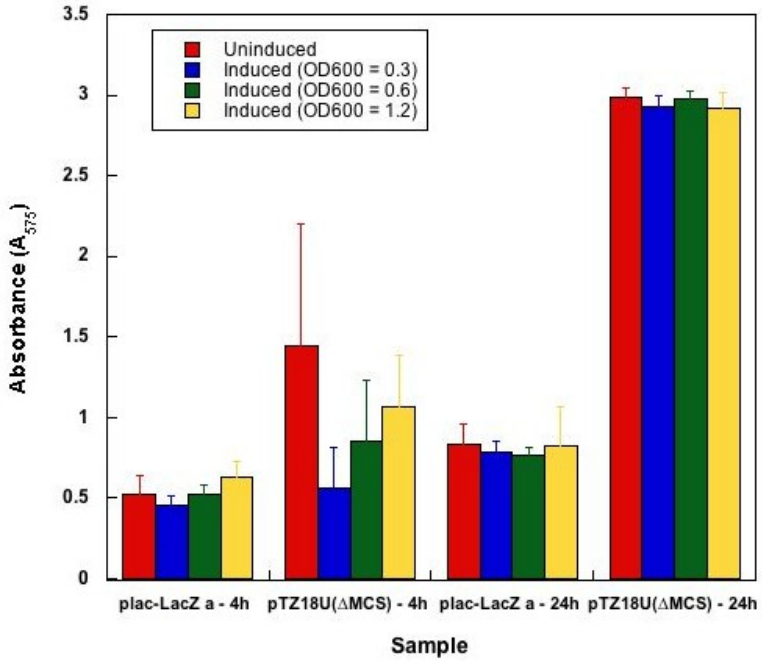


Figure V-5: β-galactosidase α-Complementation Assay with CPRG Substrate. Both the low-copy and high-copy number counter screen vectors with altered *lacZα* (plac-LacZα and pTZ18U(ΔMCS), respectively) were assayed in replicate (n=6). Cells were plated (OD_{600} = 0.3 (■), 0.6 (■), and 1.2 (■)) prior to induction with IPTG in 384-well microtiter plates. Aliquots were also removed from all samples prior to addition of IPTG as the uninduced control (■). Samples were treated with CPRG and Triton X-100 following 4 hrs and 20 hrs of growth at 37°C. The chlorophenol red product was detected spectrophotometrically (A_{575}). The average of each trial is illustrated with the error represented as standard deviation.

optimize the CPRG assay. Maximal LacZ expression was observed in both constructs following induction with a relatively low concentration of IPTG (0.5

mM). In addition, pET-LacZ expression was tested in the permeable *E. coli* strain, EC2880 (Δ*tolC, imp-,* Δ*lac*). This cell line has a deletion in the entire *lac* operon, and is therefore unsuitable for α-complementation. β-galactosidase activity in EC2880 was increased 20-50% in comparison to the corresponding conditions in TG2 cells (Figure V-6).

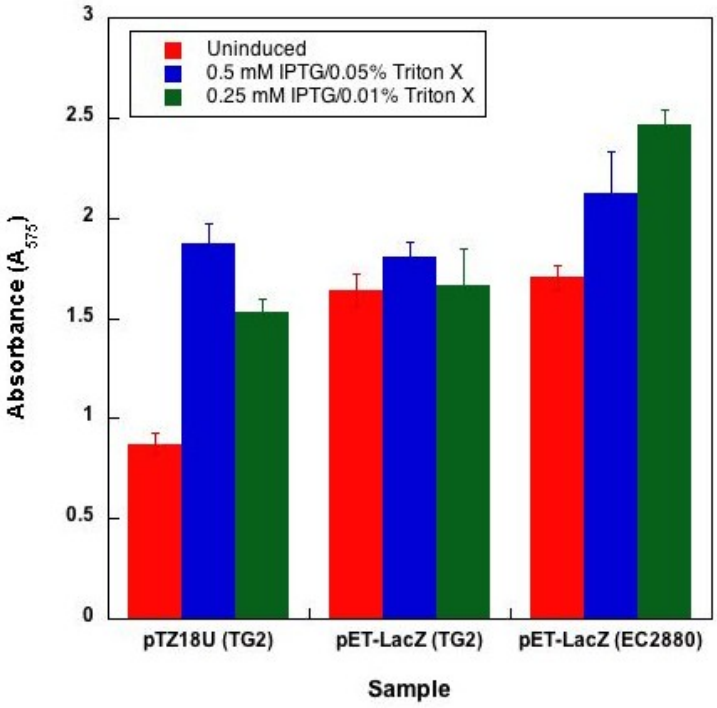


Figure V-6: Comparison of β-galactosidase α-Complementation and Expression of Full-length Reporter Gene. The high-copy number counter screen vectors with either wildtype *lacZα* (pTZ18U) or *lacZ* (pET-LacZ) were assayed in triplicate. The pET-LacZ construct was assayed in both *E. coli* TG2 and the permeable *E. coli* stain, EC2880. Cells (OD_{600} = ~1.5) were incubated for 24 hours at 37°C in 384-well microtiter plates with no IPTG (■), 0.5 mM IPTG (■) or 0.25 mM IPTG (■). The β-galactosidase assay was performed with addition of CPRG and Triton X-100. The chlorophenol red product was detected spectrophotometrically (A_{575}). The average of each trial is illustrated, where the error is represented as standard deviation.

Further optimization of the VirF reporter assay was conducted with constructs expressing the full-length LacZ protein. Two high-copy number vectors (the *tac* promoter-based vector, pMALvirF-LacZ and the L-arabinose-inducible plasmid, pBADvirF-LacZ) expressing the MalE-VirF fusion protein and the *virB* promoter-driven *lacZ* gene were tested in EC2880. β-galactosidase activity was high in the uninduced samples for each reporter following growth in microtiter plates at 37°C (Figure V-7). Expression of LacZ in the IPTG-induced pMALvirF-LacZ samples was 3-4 fold lower than the uninduced cells following 18 hours of growth. β-galactosidase activity was observed following much shorter incubation times in the pBADvirF-LacZ samples (4 hr vs. 18 hr); interestingly, identical levels of β-galactosidase activity was observed in the presence and absence of L-arabinose.

In an attempt to overcome the apparent "leaky" expression of β-galactosidase observed in the previous trials, both VirF-LacZ reporter constructs were assayed in *Shigella flexneri* BS103 (virulence plasmid-cured derivative of *S. flexneri* 2a). Cells were grown to mid-log phase and plated in varying concentrations of inducer for 3 hours at 37°C. BS103 cells exhibited no endogenous β-galactosidase activity in the presence of either IPTG or L-arabinose (Figure V-8A and B, respectively). Uninduced pMALvirF-LacZ again produced the highest level of β-galactosidase activity, although at a faster rate than was observed in *E. coli* EC2880 (Figure V-8A). Again, increasing concentration of IPTG resulted in decreased expression of LacZ in the pMALvirF reporter. LacZ production in BS103 was unaffected by the concentration of L-

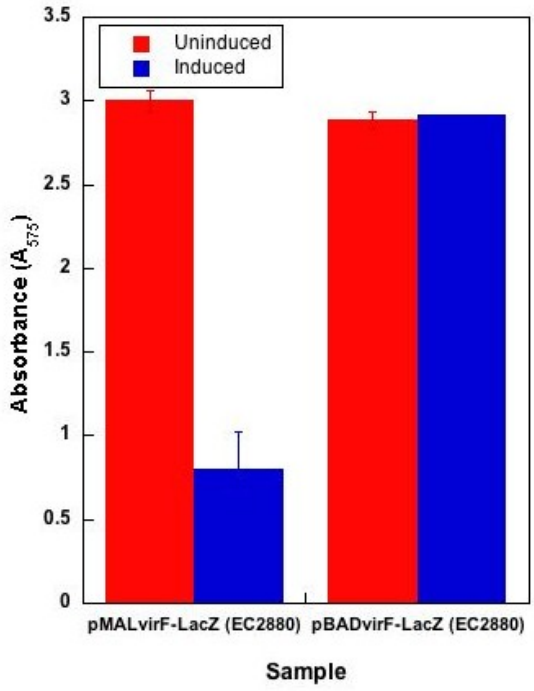


Figure V-7: VirF-β-galactosidase Reporter Assay in Permeable *E. coli*. The high-copy number vectors pMALvirF-LacZ and pBADvirF-LacZ were assayed in duplicate in EC2880. Cells (OD_{600} = ~0.7-0.8) were incubated without inducer (■) and in the presence of the highest concentration of inducer tested (■, 2.0 mM IPTG for pMAL and 0.02% L-arabinose for pBAD). The pMAL reporter was incubated for 18 hours at 37°C in 384-well microtiter plates, and the pBAD reporter for 4 hours at 37°C. The β-galactosidase assay was performed with additon of CPRG and Triton X-100, and the chlorophenol red product was detected by absorbance (A_{575}). The average of each trial is illustrated, where the error is represented by the standard deviation.

arabinose in the pBADvirF reporter construct, but maximal expression of β-galactosidase was 2-fold lower than was observed for the pMALvirF-LacZ reporter (Figure V-8B). To confirm that expression of LacZ was VirF-dependent in BS103, a mutant version (ΔvirF) of each reporter was constructed. Uninduced β-galactosidase activity was monitored from the pMALvirF-LacZ and pMAL(ΔvirF)-LacZ reporters in BS103, and VirF-driven expression of LacZ was 4-fold higher than the positive control (Figure V-8C). CPRG-mediated detection

of LacZ expression from the pMALvirF-LacZ reporter in BS103 were chosen as the optimal assay conditions for further trials.

Optimization of High-throughput VirF-LacZ Assay

The microtiter plate-based assay was optimized at the Center for Chemical Genomics (CCG, University of Michigan). Assays were performed in the absence of small molecules according to the desired protocol for the final assay. Each set of conditions was performed in replicate (n=352) to determine the consistency of cell growth and β-galactosidase activity across each well in the plates. Conditions such as growth temperature (30 vs. 37°C), total volume (60 vs. 80 µL), and concentrations of the CPRG substrate (0.25 vs. 0.50 mg/mL) and Triton X-100 detergent (0.01 vs. 0.05%) were considered in the initial troubleshooting measures. The most reproducible β-galactosidase assay was observed when cells (OD_{600} = 0.004) were incubated at 30°C for approximately 24 hours, and assayed with 0.50 mg/mL CPRG + 0.05% Triton X-100 in a 60 µL total volume (Figure V-9). Cell growth was measured (A_{600}) to ensure equivalent cell density between replicate wells (data not shown). The heat map representation illustrates the β-galactosidase activity in each well of the 384-well microtiter plate following treatment with CPRG and Triton X-100. The negative control (pMALvirF-LacZ without compound) was plated in the first two columns (n=32), and the average absorbance was recorded (A_{570} = 2.5). The positive control (pMAL(ΔvirF)-LacZ) was plated in the last two columns (n=32), with an average A_{570} = 0.72. The “test set” (pMALvirF-LacZ in this case) was plated in

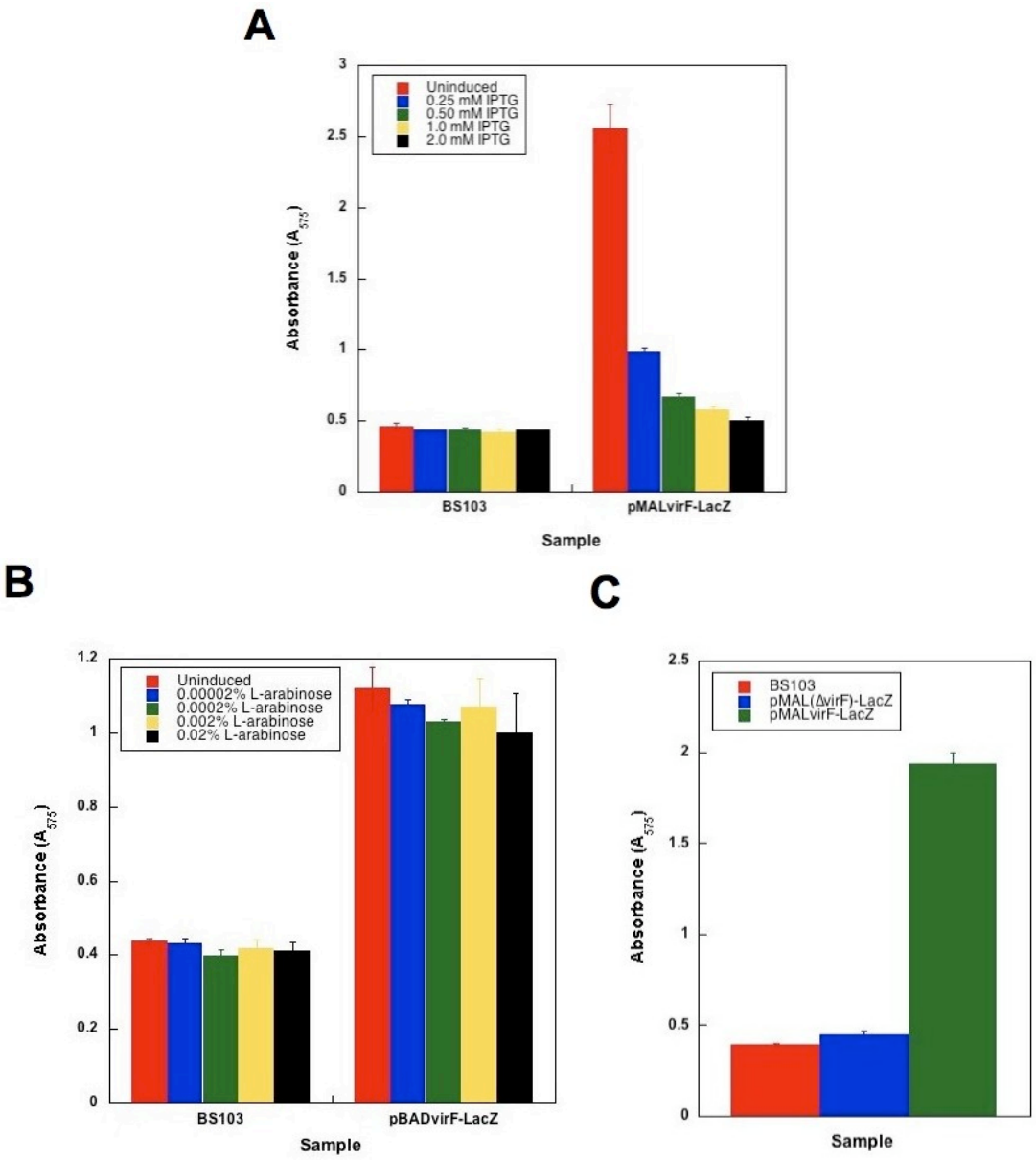


Figure V-8: VirF-β-galactosidase Reporter Assay in Avirulent *Shigella flexneri*. A) pMALvirF-LacZ in BS103. All samples were grown to OD_{600} = ~0.6-0.7 and plated with varying concentrations of inducer in duplicate: uninduced (■), 0.25 mM (■), 0.50 mM (■), 1.0 mM (■), and 2.0 mM (■) IPTG. Samples were treated with CPRG and Triton X-100 following a 3 h induction period, and the chlorophenol red product was detected by absorbance (A_{575}). B) pBADvirF-LacZ in BS103 were assayed as above. The concentration of L-arabinose was varied as follows: uninduced (■), 0.00002% (■), 0.0002% (■), 0.002% (■), and 0.02% (■) L-arabinose. C) β-galactosidase activity in BS103 with pMALvirF-LacZ or pMAL(ΔvirF)-LacZ. Samples were grown to a final OD_{600} = ~0.5 and plated in the absence of IPTG in replicate (n=4). All samples were assayed as described above.

the internal 320 wells, and β-galactosidase activity was measured in comparison to the controls by % inhibition (100% inhibition set to the A_{570} reading for the positive control). There was some variation in growth between individual wells, but the overall Z' factor determined was reasonable (Z' = 0.59). In addition, a plot of the raw absorbance values indicated that most of the pMALvirF-LacZ replicates achieved ≥ 3 SD (standard deviations) above the Δ*virF* positive control.

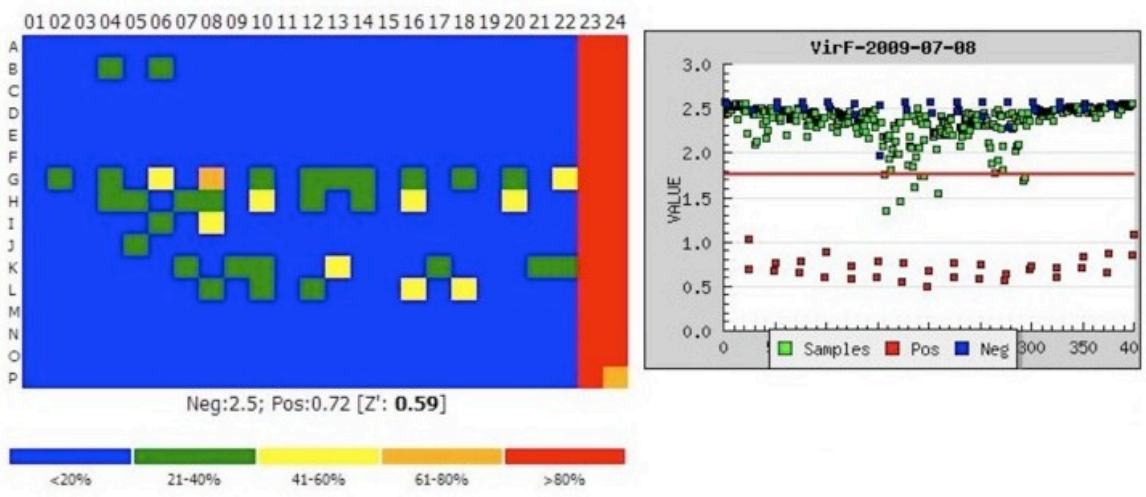


Figure V-9: Optimization of VirF-β-galactosidase Reporter Assay in 384-well Microtiter Plate. *S. flexneri* BS103 transformed with either pMALvirF-LacZ or pMAL(ΔvirF)-LacZ were plated in a 30 µL volume at a final OD_{600} = 0.004 (columns 1-22: pMALvirF-LacZ, columns 23-24: pMAL(ΔvirF)-LacZ). The microtiter plate was incubated at 30°C overnight. β-galactosidase activity was measured (A_{570}) upon addition of CPRG and Triton X-100. The calculated Z' factor = 0.59. The data is visualized both as a heatmap of assay consistency and by the measured absorbance value for each well.

The primary HTS screen of the VirF-LacZ cell-based assay was run with several small molecule libraries, including MicroSource Spectrum 2000, the Chemical Diversity set, Maybridge Hit-Finder, and the NCC BioFocus library.

Approximately 42,000 compounds were tested in total, and the raw absorbance data is shown in Figure V-10. Cell growth was monitored (A_{600}, Figure V-10A) prior to addition of the CPRG substrate/detergent mixture. Each small molecule was tested in a single well (n=1) in a 384-well microtiter plate, and ~1270 compounds were selected as potential hits based on the following selection criteria (normalized to the controls of each individual plate) (Figure V-10B): ≥ 55% VirF inhibition by plate (A_{570} < 1.4), with 100% inhibition defined as the mean A_{570} measurement for the positive control replicates for each plate (A_{570} ≅ 0.7-1.0). Compounds with antibiotic activity of ≥ 20% growth inhibition were also removed from consideration. The 1273 initial hits were re-assayed using the primary screen conditions in triplicate to confirm activity (Figure V-11A).

In addition, these compounds were counter-screened in triplicate against purified β-galactosidase, *in vitro*, to filter out any false positives that directly inhibit the reporter (Figure V-11B). Four compounds were found to inhibit β-galactosidase *in vitro* (Table V-2), three with modest inhibition (~40%) and one that demonstrated significant inhibition of the purified enzyme (92%, CCG-28697). Three compounds (CCG-28697, 28450, and 28967) share a common tricyclic structure with a bridging piperdine-sulfonyl moiety, derivatized with various aromatic functionalities (Table V-2). The structures are distinct from galactose, and it is unclear how these compounds inhibit β-galactosidase activity.

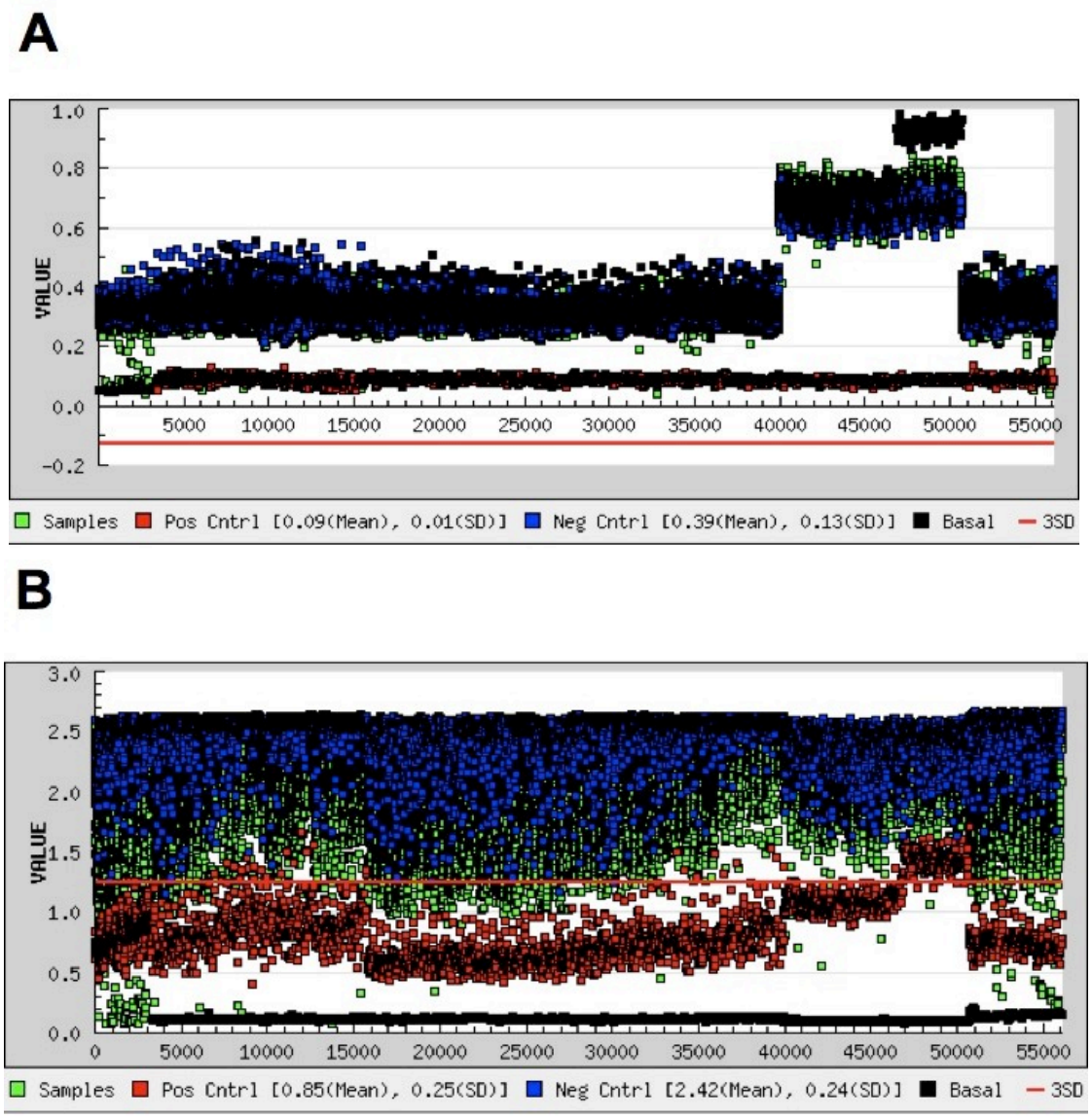


Figure V-10: VirF-β-galactosidase Reporter Assay Primary Screen. *S. flexneri* BS103 harboring pMALvirF-LacZ or pMAL(ΔvirF)-LacZ were plated in replicate (n=32) to a final OD_{600} = 0.004 and grown at 30°C overnight. Small molecules (10 μM) were added by pintool (320 compounds/plate, n=1). Approximately 42,000 small molecules were tested in the primary screen. A) Cell growth (OD_{600}) was monitored prior to addition of CPRG/detergent mixture as a control for antibacterial activity. B) β-galactosidase activity was measured (A_{570}) upon addition of CPRG and Triton X-100.

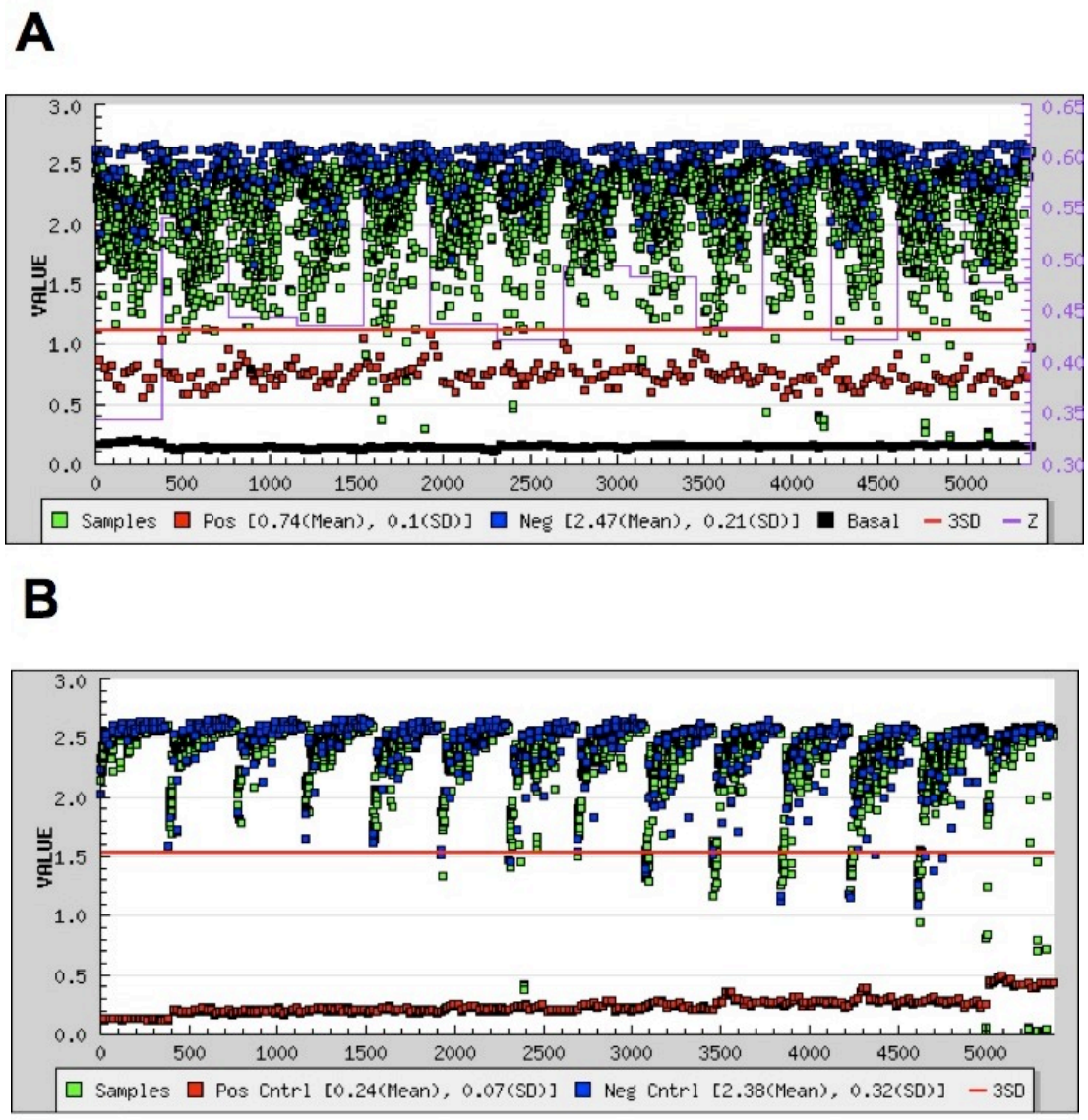
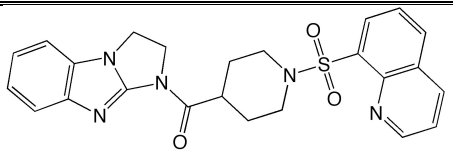
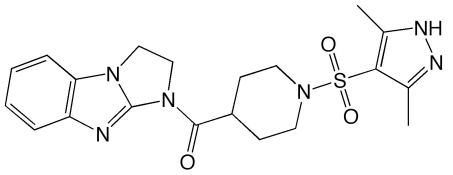
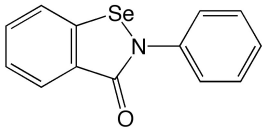
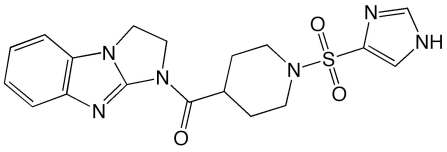


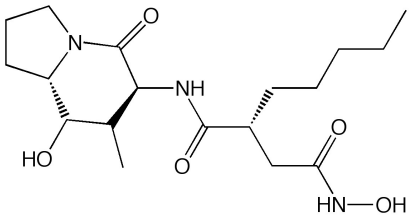
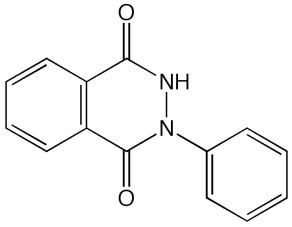
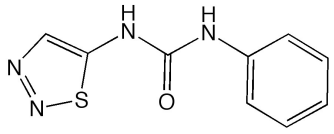
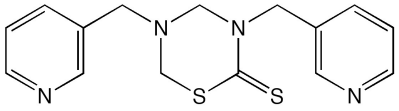
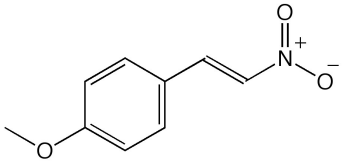
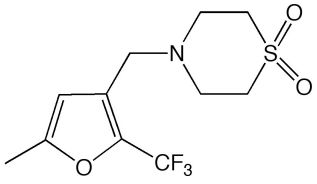
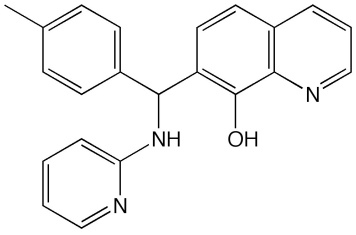
Figure V-11: VirF-β-galactosidase Reporter Assay Confirmation Screen. A) Compounds identified as potential hits in the primary screen (1273 small molecules, 10 μM) were re-assayed to confirm activity in triplicate. The assay was carried out as described in the text. B) Identical plates were prepared to counterscreen with purified β-galactosidase (25 nM) to monitor direct inhibition of the reporter *in vitro*. The positive control (enzyme without compound) and negative control (Z-buffer) were plated in replicate (n=32).

Table V-2: Confirmed Inhibitors of Purified β-galactosidase *in vitro*.

Compound Name	Library	Percent Inhibition (Ave. by Plate)	Structure
CCG-28697	ChemDiv	92%	
CCG-28450	ChemDiv	40%	
CCG-39161	MicroSource	40%	
CCG-28967	ChemDiv	35%	

Following further selection criteria for optimized hit-to-lead scaffolds (i.e., molecular weight < 600, no reactive or toxic functionalities) a test set of 238 compounds was chosen for dose response studies. Each compound was tested in replicate (n=4) in serial dilutions ranging in concentration from 100 - 1.5 μM. The data were plotted (concentration versus absorbance) fit by non-linear regression using an equation for sigmoidal kinetics, and the IC_{50} values were calculated using Kaleidagraph (Synergy Software). From this assay, 7 compounds were confirmed by dose response analysis, with typical IC_{50} values ranging between 1-90 μM (Table V-3). Representative data for the dose response analysis is shown in Figure V-12.

Table V-3: Dose Response Analysis for VirF-β-galactosidase Reporter Assay.

Compound Name	Library	IC_{50} (μM)	Structure
CCG-38543	MicroSource	1.0	
CCG-25354	ChemDiv	2.7	
CCG-24904	ChemDiv	4.5	
CCG-50009	Maybridge	10	
CCG-55215	Maybridge	12	
CCG-42010	Maybridge	43	
CCG-21496	ChemDiv	89	

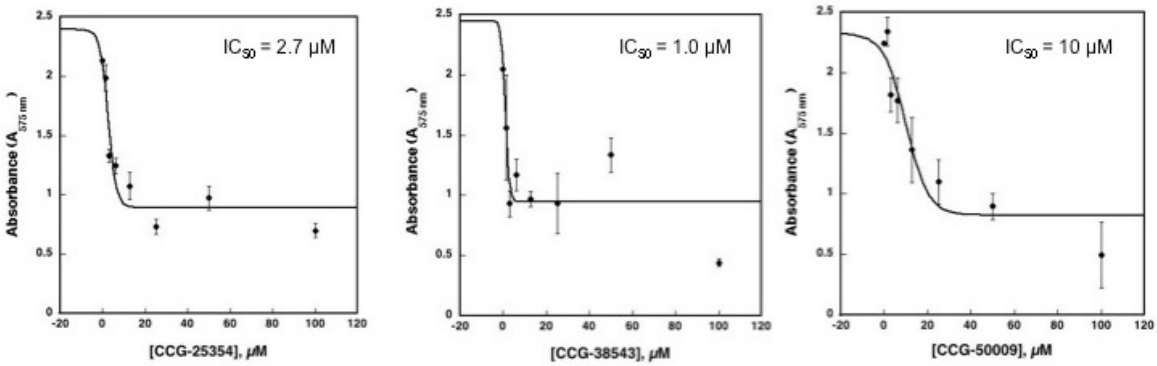


Figure V-12: VirF-β-galactosidase Reporter Assay Dose Response Analysis. Three compounds fit by non-linear regression for sigmoidal kinetics. Each compound was confirmed in replicate (n=4).

Discussion

In considering a robust, cell-based assay to screen for inhibitors against VirF activity, β-galactosidase was initially chosen as a reporter for its history as a molecular biology tool and its enzymatic stability [10]. With a wide variety of substrates described in the literature, there are several potential options to suit a range of β-galactosidase assay protocols (Figure V-13). The most common substrate for measuring LacZ expression on solid media is X-gal (5-bromo-4-chloro-3-indolyl β-D-galactopyranoside), however detection of the insoluble blue product is less amenable to liquid culture. The classical substrate described by J. H. Miller for β-galactosidase enzyme kinetics is *o*-nitrophenyl β-D-galactopyranoside (ONPG) [10]. The ONP product of hydrolysis is yellow in color and detected by absorbance (A_{420}). This spectrophotometrically active substrate is best suited to *in vitro* reactions because the signal to noise ratio is high in colorless buffer. For use in cell-based assays, however, detection of ONP is

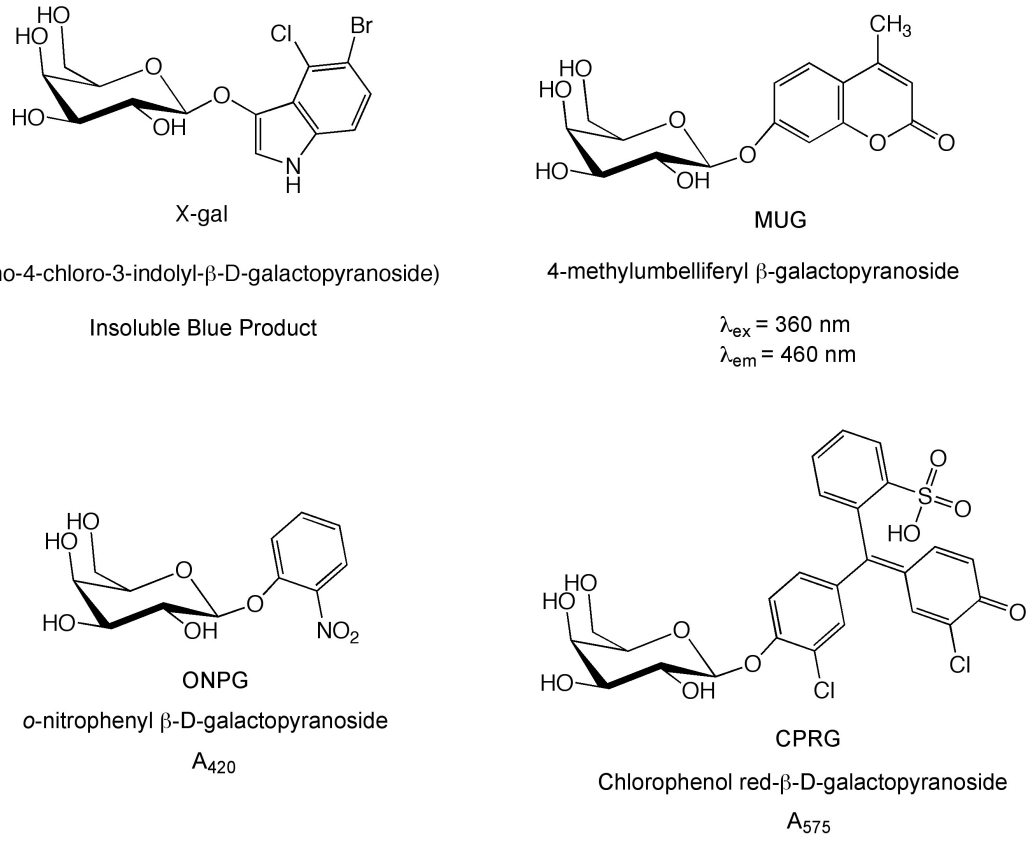


Figure V-13: β-galactosidase Substrates. Chemical name and detection method for the reaction product is listed.

masked by the inherent absorbance in the liquid media. The fluorescent substrate MUG was previously described for use in a whole-cell assay for β-galactosidase activity [11]. Without significant dilution of the starting cell culture, there is a concern for background fluorescence from the media. Finally, CPRG (chlorophenol red β-D-galactopyranoside) has also been described for cell-based applications, and this substrate is particularly sensitive to low concentrations of β-galactosidase [12].

In addition to choosing the optimal substrate, LacZ enzyme activity can be monitored from expression of the full-length protein or by α-complementation. Initially we considered use of the *lacZ* α-compliment because this short peptide

sequence is readily available in most plasmids as a tool for blue/white selection of positive transformants. LacZ α-complementation is achieved through expression of the α-peptide in cells with the genotype *lacZ*Δ*M15*, which indicates that only the LacZ Ω-peptide is present. Cell lines with a Δ*lac* genotype must be transformed with a plasmid carrying the entire *lacZ* gene sequence to exhibit β-galactosidase activity. For our purposes and ease in future subcloning steps, the multiple cloning site (MCS) of the *lacZ* α-peptide in pTZ18U was replaced by a short cassette that did not disrupt the reading frame. The corresponding *lacZ* α-peptide contains an altered N-terminus, with a deletion in 10 amino acids and several major amino acid substitutions (e.g. P19R, D21T, C27A, H29E). When transformed into TG2 cells and plated with X-gal and IPTG, the colonies exhibited blue/white selection, indicating that although some pronounced sequence alterations were made, the resultant LacZ protein is still functional.

The *lacZ*(ΔMCS) fragment was amplified by PCR, and subcloned initially into the low/medium copy-number vector pACYC184. Because VirF-mediated activation of the *virB* promoter is quantified by resultant β-galactosidase activity, we hypothesized that detection of LacZ might be more sensitive if the concentration of the *lacZα* gene was lower in cells. Assays were first conducted with the constituitive *lac* promoter-driven expression vectors in an attempt to optimize conditions for the ultimate VirF reporter assay (Figure V-4A). β-galactosidase activity was quantified in a modified Miller method using a cell-based high-throughput assay with the LacZ substrate, MUG [11]. The Vidal-Aroca method monitored β-galactosidase activity in black/clear-bottom microtiter

plates, however fluorescence measurements were more consistent in black opaque plates. For this reason, the preliminary absorbance readings were omitted, and the data was reported in fluorescence units as opposed to Miller Units. One difficulty with the MUG assay was the increased background due to inherent fluorescence in the media (Figure V-4B). This increase is likely due to the presence of riboflavin in the rich media, which fluoresces at approximately the same wavelengths (λ_{ex} = 350 nm, λ_{em} = 450 nm). Some success in reducing the signal to background ratio was achieved by conducting the assay with p*lac*-LacZα in minimal media, but it seems unlikely that these conditions would translate well to the VirF reporter assay (see Chapter IV: Troubleshooting VirF Expression). In addition, optimal β-galactosidase activity results were observed following induction in shaking, independent cultures. This technique is suitable for a small volume of samples but would never be amenable to development of a high-throughput assay. Ultimately, use of the MUG substrate was abandoned due also to the non-specific hydrolysis observed in plasmid-free TG2 cells (data not shown).

CPRG was the next β-galactosidase substrate selected for assay optimization due to the sensitivity of CPRG to low LacZ concentrations. In addition, the yellow substrate turns a deep red upon hydrolysis by β-galactosidase, and accumulation of the chlorophenol red product is easily detected by absorbance. One potential concern with use of CPRG was the optical density of the cells necessary for conducting the VirF assay. The whole-cell assay could require long growth periods to produce enough VirF protein that

could in turn activate expression of a reasonable amount of β-galactosidase. The optical density is typically measured at A_{600}, and a high cell content could complicate the measurement of chlorophenol red at A_{575}. Because high background had also been an issue with MUG, it was important to develop this method further with the *lac* promoter-driven constructs.

Trials with CPRG were conducted to monitor LacZ α-complementation in both pACYC184 and pTZ18U to address the issue of plasmid copy number (Figure V-5). The α-complementation assay with pTZ18U(ΔMCS) produced higher absorbance measurements in the CRPG assay than the low copy vector, p*lac*-LacZα (3-6 fold). However, the variability between replicates was also increased in the high copy-number vector. To determine whether this observation was a result of the alteration in *lacZα*, wildtype pTZ18U was also assayed in TG2 cells. The replicates were much more consistent in the unaltered α-complementation assay (Figure V-6). Expression of the full-length *lacZ* gene from a high copy-number vector, pET-LacZ (pET28a derivative), was also monitored for comparison between α-complementation and expression of the native reporter sequence (Figure V-6). The most successful expression of LacZ in a *lac* promoter-based vector was demonstrated in the permeable *E. coli* strain, EC2880. This cell line was commonly used in high-throughput screening endeavors at Pfizer. The Δ*tolC* mutation results in increased drug permeability and allows for study of molecules that might otherwise be removed by the drug efflux pump. With these results taken together, we made several conclusions in moving forward with the VirF reporter assay: 1) measuring β-galactosidase

activity with α-complementation is not suitable for detailed assay measurements, 2) expression of VirF and LacZ in a single, high-copy vector would be optimal, and 3) detection of β-galactosidase activity by CPRG is the most sensitive in liquid culture.

Following construction of the all-encompassing VirF reporter plasmids, pMALvirF-LacZ and pBADvirF-LacZ, each construct was assayed in EC2880. β-galactosidase activity was very high in the uninduced samples, but activity was reduced in pMALvirF-LacZ supplemented with 2 mM IPTG (Figure V-7). Interestingly, inhibition of VirF in the presence of IPTG has been observed previously [13]. Standard induction with 1 mM IPTG resulted in growth inhibition in *Shigella*, and VirF expression was achieved only through induction with submillimolar concentrations (50-100 μM IPTG). One potential explanation for this uninduced β-galactosidase activity is the result of leaky VirF expression. Another concern is the possibility that the *virB* promoter controlling expression of the *lacZ* gene is being activated by an *E. coli* transcription factor. A report published by C. Dorman and colleagues detailed the cross-reactivity of the enterotoxigenic *E. coli* AraC-type transcriptional activator, Rns, which was shown to activate transcription of VirF-specific promoter sequences [14]. It is possible that we are detecting Rns-activated expression of β-galactosidase in our *E. coli* cell-based assay, as opposed to leaky expression of the VirF protein.

Because there is no other known transcriptional activator in *Shigella* that can activate VirF promoters, the VirF reporter assay was optimized in an avirulent strain of *Shigella flexneri*, BS103 [15]. BS103 is a derivative of *S.*

flexneri 2a that has been cured of the 220 kb virulence plasmid, and hence lacks endogenous virulence gene expression (including *virF*, *virB*, and *virR* (*hns*)). As with *E. coli*, uninduced β-galactosidase activity was observed in both pMALvirF-LacZ and pBADvirF-LacZ (Figure V-8A and B, respectively). With no other activator capable of initiating transcription of the *virB-lacZ* open reading frame, the resultant activity was hypothesized to be the result of leaky VirF expression. With no known small molecule inhibitor of VirF, a deletion in the *malE-virF* open reading frame was constructed in each reporter plasmid to monitor any background LacZ expression not related to VirF-specific activation. This construct was in every other aspect identical to the reporters discussed previously. Assays were conducted with the pMAL(ΔvirF)-LacZ reporter due to the increased β-galactosidase activity observed in the *tac* promoter-based system. We confirmed that LacZ expression in BS103 is the direct result of VirF activation of the *virB* promoter, with the Δ*virF* reporter responding similarly to the Δ*lac* BS103 cell control (Figure V-8C).

With optimized conditions for the reporter construct and positive control, β-galactosidase substrate for the cell-based assay, and *Shigella* cell line in place, the high-throughput assay methods were defined in conjunction with the Center for Chemical Genomics (University of Michigan). One primary difference in HTS assay setup in comparison to the lab protocol described previously is the initial growth phase. Cells had previously been grown in batch culture with shaking to the optimal induction OD_{600}, then cells were plated from this common stock for induction in microtiter plates. Reproducibility in the new assay setup was critical

to establish that cell growth and corresponding β-galactosidase activity would be uniform. Another benefit to monitoring assay conditions at the CCG is the narrow band pass filter used in the PHERAstar plate reader, which has an increased sensitivity in absorbance measurements in comparison to the monochromator detection of the Molecular Devices plate reader. Conditions for optimal cell growth and β-galactosidase activity were identified in 384-well microtiter plate format, and a reasonable Z' factor = 0.59 was achieved (Figure V-9).

Although enzymatic activity was relatively consistent across the plate, there was an issue of evaporation around the perimeter of each microtiter plate. These so called "edge effects" were also observed in other cell-based assays performed at the CCG. In these wells, although the OD_{600} reading was not altered significantly, the total volume was reduced by 5-10 µL. This phenomenon was more pronounced following overnight growth at 37°C. Because the virulence plasmid has been removed from the BS103 cells, there was no concern of temperature-regulated VirF inhibition with growth at lower temperatures, so HTS was performed at 30°C. Another issue of consistency observed in the cell-based reporter assay was the length of incubation time. Typically cells were plated in the presence of compound for approximately 24 hours prior to addition of the substrate/detergent mixture, and the relatively consistent OD_{600} readings indicate that cell growth was uniform across the primary assay (Figure V-10A). However, one set of plates was measured following ~30 hours of incubation (Cmpds. 40,000-50,000), and both the OD_{600} measurement and A_{570} values for these plates have a skewed background for β-galactosidase activity (Figures V-

10A and B, respectively). Although there was discrepancy in some cases between wells and plates, the internal controls for each plate were used to normalize the data. Overall the "noisy" data is reflected in a lower Z' factor ≅ 0.4, but we felt keeping all the data was most beneficial to our understanding of small molecule inhibition of VirF in the reporter assay.

Approximately 3% of the compounds tested in the primary screen demonstrated modest to significant inhibition of the VirF-LacZ assay. These compounds were confirmed in triplicate according to the primary assay protocol. The confirmation screen was more consistent, although there were a few new compounds that exhibited antibacterial activity previously undetected (Figure V-11A). As a preliminary counter screen, purified β-galactosidase was pre-incubated with the same small molecules *in vitro*. Most compounds tested in the confirmation screen did not inhibit the reporter directly (Figure V-11B, Table V-2). One concern is this counter screen does not address the issue of non VirF-specific activity. The best counter screen to address off-target activity would be another cell-based system where LacZ expression was activated by the endogenous *Shigella* transcription complex. We had initially considered this with the *lac* promoter-driven reporter plasmid (p*lac*-LacZα). However, *Shigella* is a Δ*lac* strain, and thus transformation with a *lac*-driven reporter would be ineffective. One potential candidate is the endogenous *trp* promoter.

The confirmation hits were triaged further, taking into account the "drugability" of the potential lead molecules. Ultimately a set of 238 compounds were selected for dose response analysis. Although most compounds did not

demonstrate dose-dependent inhibition of the VirF-LacZ assay, 7 small molecules were identified, with IC_{50} values in the low-mid micromolar range (Table V-3). The data were fit by non-linear regression for sigmoidal kinetics, and the reported pIC_{50} values were converted to IC_{50} values. Interestingly, several compounds identified were also characterized as "active" in other bacterial cell-based assays targeting transcription factors. The 7 confirmed compounds identified in the VirF-β-galactosidase assay have each been tested in ≥ 35 individual assays, including screens against prokaryotic and eukaryotic targets *in vitro* and *in vivo*. In general, the compounds have demonstrated little activity against eukaryotic targets. The similarity in effects against assays with structurally distinct proteins performing similar functions (*i.e.,* DNA-binding, transcriptional activation) could suggest that these small molecules target a conserved function of bacterial transcription factors. Further studies are needed to confirm the specificity of these small molecules.

Conclusions

In collaboration with the University of Michigan, Center for Chemical Genomics (CCG), we have developed a cell-based bacterial transcription assay amenable for high throughput screening (HTS) of large compound libraries against VirF from *Shigella flexneri*. Initial HTS results screening ~42,000 small molecules in the VirF-LacZ reporter assay identified 238 potential hits with favorable medicinal chemistry properties (i.e., molecular weight < 600, no reactive or toxic functionalities, logP < 7). Dose response data for 7 compounds revealed IC_{50} values in the micromolar range (1-90 μM). Future work will be

conducted with several secondary assays (i.e., *in vitro* DNA-binding assay and *in vitro* transcription assay) to probe the mechanism of inhibition for these “hits”. Work will also be conducted with other AraC-type proteins to identify scaffolds with broad-spectrum activity in this family of bacterial transcriptional activators.

Notes to Chapter V

We gratefully acknowledge Martha Larson, Thomas McQuade, and Dr. Paul Kirchoff with the University of Michigan Center for Chemical Genomics for their assistance with the VirF high-throughput assay development and triage. We also thank Prof. Anthony Maurelli for the avirulent *Shigella* strain, BS103. I would also like to thank Anthony Emanuele, a former undergraduate student who worked closely with me on the development of the various VirF reporter constructs.

Abbreviations used: CPRG, chlorophenol red β-D-galactopyranoside; dNTP, deoxyribonucleotide triphosphate; CIP, calf-intestinal alkaline phosphatase; IPTG, isopropyl β-D-thiogalactoside; X-gal, 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside; MCS, multiple cloning site; DTT, dithiothreitol; Tris•HCl, tris(hydroxymethyl)aminomethane hydrochloride; PEG, polyethylene glycol; PCR, polymerase chain reaction; RBS, ribosome binding site; MUG, 4-methylumbelliferyl β-D-galactopyranoside; ONPG, *o*-nitrophenyl β-D-galactopyranoside; DMSO, dimethyl sulfoxide; AC_{50}, half-maximal activity concentration; CCG, Center for Chemical Genomics; HTS, high-throughput screening.

1. Gallegos, M.T., et al., *Arac/XylS family of transcriptional regulators.* Microbiology & Molecular Biology Reviews, 1997. **61**(4): p. 393-410.

2. Soisson, S.M., et al., *Structural basis for ligand-regulated oligomerization of AraC.* Science, 1997. **276**(5311): p. 421-425.

3. Falconi, M., et al., *Thermoregulation of Shigella and Escherichia coli EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of virF promoter to transcriptional repressor H-NS.* EMBO Journal, 1998. **17**(23): p. 7033-43.

4. Martin, R.G. and J.L. Rosner, *The AraC transcriptional activators.* Current Opinion in Microbiology, 2001. **4**(2): p. 132-137.

5. Schleif, R., *AraC protein: a love-hate relationship.* Bioessays, 2003. **25**(3): p. 274-82.

6. Tobe, T., et al., *Transcriptional control of the invasion regulatory gene virB of Shigella flexneri: activation by virF and repression by H-NS.* Journal of Bacteriology, 1993. **175**(19): p. 6142-9.

7. Porter, M.E. and C.J. Dorman, *In vivo DNA-binding and oligomerization properties of the Shigella flexneri AraC-like transcriptional regulator VirF as identified by random and site-specific mutagenesis.* Journal of Bacteriology, 2002. **184**(2): p. 531-9.

8. Kwon, H.J., et al., *Crystal structure of the Escherichia coli Rob transcription factor in complex with DNA.* Nature Structural Biology, 2000. **7**(5): p. 424-30.

9. Bowser, T.E., et al., *Novel anti-infection agents: small-molecule inhibitors of bacterial transcription factors.* Bioorganic & Medicinal Chemistry Letters, 2007. **17**(20): p. 5652-5.

10. Miller, J.H., *Experiments in Molecular Genetics*. 1972, Cold Spring Harbor, NY: CSH Laboratory Press.

11. Vidal-Aroca, F., et al., *One-step high-throughput assay for quantitative detection of beta-galactosidase activity in intact gram-negative bacteria, yeast, and mammalian cells.* Biotechniques, 2006. **40**(4): p. 433-4.

12. Eustice, D.C., et al., *A Sensitive Method for the Detection of Beta-Galactosidase in Transfected Mammalian-Cells.* Biotechniques, 1991. **11**(6): p. 739-&.

13. Durand, J.M.B., et al., *Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of Shigella flexneri: the expression of the virF gene.* Molecular Microbiology, 2000. **35**(4): p. 924-935.

14. Porter, M.E., S.G. Smith, and C.J. Dorman, *Two highly related regulatory proteins, Shigella flexneri VirF and enterotoxigenic Escherichia coli Rns, have common and distinct regulatory properties.* FEMS Microbiology Letters, 1998. **162**(2): p. 303-9.

15. Maurelli, A.T., B. Blackmon, and R. Curtiss III, *Loss of pigmentation in Shigella flexneri 2a is correlated with loss of virulence and virulence associated plasmid.* Infection and Immunity, 1984. **43**: p. 397-401.

CHAPTER VI

Summary

Shigella flexneri, a causative agent of bacillary dysentery, is an enteropathogen that invades the intestinal mucosa. A food- and water-borne pathogen, *Shigella* infections are most commonly found in third-world countries where sanitation standards are lower than in the more developed nations [1]. Increasing emergence of antibiotic-resistant strains of *Shigella* heightens the need for identification of novel antibacterial targets for a directed drug discovery approach in the treatment of shigellosis. One facet of *Shigella* infection is the expression of key virulence factors (*i.e.,* VirF, VirB, TTSS), found on a large molecular-weight virulence plasmid, that promote cell contact and expression of all downstream virulence genes [2]. The master positive regulator of the virulence plasmid is VirF, an AraC-type transcriptional activator, which functions primarily to activate the secondary virulence regulator, VirB. Mutant *Shigella flexneri* (Δ*virF*) strains have been characterized as avirulent, demonstrating the importance of this key virulence factor in the pathogenicity cascade.

Many cellular factors have been implicated in the transcription of *virF* mRNA (*i.e.,* DNA superhelicity, pH, temperature) and translation of the VirF protein (*i.e.,* concentration of oxygen, iron, amino acids) [3-5]. Interestingly, VirF protein levels in an *S. flexneri* (Δ*vacC*) strain were reduced to ~40% in comparison to the concentration of VirF in wild-type *Shigella* [6]. VacC is the

Shigella homologue of tRNA-guanine transglycosylase (TGT), which catalyzes the incorporation of the non-canonical nucleoside queuosine (Q) into four substrate tRNAs ($tRNA^{Asn,\ Asp,\ His,\ Tyr}$). This relatively subtle decrease in VirF protein expression resulted in a decreased ability for the bacteria to invade host cells in a hemolytic assay. It is conceivable that a modest VirF inhibitor could ultimately become a potent antibiotic against *Shigella*. We proposed targeting the VirF protein as a novel antibacterial target in *Shigella flexneri*. The purpose of this dissertation is to address both the role of TGT and the queuine modification in expression of the VirF protein and identify potential inhibitors of VirF through development of a high-throughput reporter assay.

We hypothesize that TGT modulates VirF expression with queuine modification at both the level of tRNA and the *virF* mRNA. We report that the modification state of tRNA with queuine and the Q-cognate codon usage (NAU versus NAC codons) in the target mRNA alters the rate of protein expression. Using GFP (green-fluorescent protein) as a model protein, we observe that expression of *gfp* mRNA with relatively equivalent queuine-cognate codon usage results in similar rates of translation with either tRNA-G_{34} or -Q_{34}. In every case, the rate of translation was slower in the mutant *E. coli* K12(Δ*tgt*) in comparison to wild-type K12. In addition, the rate of GFP translation was slower for every codon-biased construct in comparison to GFP(wt), with the exception of GFP(U) in K12. The rate of translation from *gfp* mRNA with a significant NAU-bias is increased in wild-type *E. coli* expressing tRNA-Q_{34}, while the rate of translation from *gfp* mRNA with pronounced NAC-bias is decreased in *E. coli* (Δ*tgt*) with

tRNA-G_{34}. *VirF* mRNA contains an overall bias of 80% for the NAU Q-cognate codons, and it is conceivable that the subtle decrease in VirF protein expression in the *S. flexneri* (Δ*vacC*) mutant could result from the absence of queuine-modified tRNA.

In addition, we report that the *virF* mRNA is itself a substrate for *E. coli* TGT *in vitro* [7]. TGT has a rather minimal recognition element: a U-G-U sequence in a hairpin loop structure. In addition to tRNA, the enzyme can recognize and modify other substrates with this motif, including a minihelical hairpin and a dimeric form of $tRNA^{Tyr}$ [8]. We observe that TGT catalyzes the incorporation of queuine into *virF* mRNA with K_M = 1.8 μM [7]. A minihelical *virF* substrate (corresponding to nucleotides 410-433) was modified by TGT with K_M = 0.87, but a *virF* mRNA(G_{421}A) point mutant was not recognized as a substrate by TGT. Taken together, these results suggest that the *virF* mRNA is site-specifically modified by TGT *in vitro*, exchanging guanosine 421 (G_{421}) with queuine. Direct recognition of *virF* mRNA could help explain the observation that the VirF protein concentration was specifically decreased in the *Shigella flexneri* (Δ*vacC*) mutant. Interestingly, this is the first observation of a modified nucleoside incorporated within the open reading frame of an mRNA.

A wide variety of techniques were employed to express recombinant VirF, including expression of several fusion tags, co-expression with molecular chaperones, decreased expression temperatures, and *in vitro* translation from both mRNA and DNA starting materials. The isolation of soluble, full-length VirF was ultimately successful only when expressed as a MalE-fusion, as described

previously [3]. Two fragments of the VirF protein were also expressed in *E. coli*, the VirF(29-263) protein with a truncation in the first 28 amino acids and VirF(154-263) containing only the DNA-binding domain. The expression and purification of soluble VirF(154-263) was successful in *E. coli*, and the truncated protein was shown to bind linear fragments of the *virB* promoter *in vitro* (Biswas and Tsodikov, unpublished). The level of VirF(29-263) expression was high in cells, but the product was predominantly insoluble. Ultimately, soluble protein was recovered from a mutant VirF(29-263, V141I) construct when co-expressed with the molecular chaperones DnaK, DnaJ, and GrpE. This point mutant (translated from the $G_{421}A$ mutant *virF* mRNA) helps to illustrate two key factors regarding the VirF protein: 1) the first 28 amino acids partially contribute to VirF insolubility, and 2) the increased solubility of this conservative mutant fragment could suggest a potential relevance to the TGT/queuine modification in the *virF* mRNA.

Because expression of purified, un-tagged VirF was unsuccessful, we were unable to gain structural information by x-ray crystallography or other biophysical methods. Instead, we focused on development of a functional assay to monitor VirF activity in the presence of potential small molecule inhibitors. A plasmid was constructed expressing the β-galactosidase reporter gene fused to the *virB* promoter, the downstream DNA target of VirF, and the *tac* promoter-controlled MalE-VirF fusion. We report VirF-specific activation of β-galactosidase activity in the virulence plasmid-cured strain *Shigella flexneri* BS103. The positive control for VirF inhibition was measured by expression from a plasmid

with a deletion in the *malE-virF* open reading frame. Approximately 42,000 small molecules were tested in a high-throughput, cell-based assay measuring β-galactosidase activity with the colorometric substrate, CPRG (chlorophenol red β-D-galactopyranoside) [9]. We have identified 7 small molecules that inhibit the VirF – β-galactosidase reporter assay in a dose-dependent manner (typical IC_{50} values between 1-100 μM). Further studies are pending to address the mechanism of action for each compound.

In conclusion, we have identified two potential avenues for TGT modulation of VirF expression, acting at the level of tRNA and *virF* mRNA modification. We have also developed a high-throughput, cell-based assay to measure transcriptional activation by the VirF protein in *Shigella*. We are working to identify small molecule inhibitors that may be transformed into new antibiotics in the treatment of shigellosis. These experiments will hopefully lead to an increased understanding of VirF from *Shigella flexneri* and ultimately this complex class of AraC-type transcriptional activators.

Notes to Chapter VI

1. World-Health-Organization, *State of the Art of New Vaccines: Research & Development*. 2003, Initiative for Vaccine Research, World Health Organization: Geneva.

2. Dorman, C.J., *The Virulence Plasmids of Shigella flexneri*, in *Microbial Megaplasmids*. 2009, Springer: Berlin / Heidelberg. p. 151-170.

3. Tobe, T., et al., *Transcriptional control of the invasion regulatory gene virB of Shigella flexneri: activation by virF and repression by H-NS.* Journal of Bacteriology, 1993. **175**(19): p. 6142-9.

4. Tobe, T., M. Yoshikawa, and C. Sasakawa, *Thermoregulation of virB transcription in Shigella flexneri by sensing of changes in local DNA superhelicity.* Journal of Bacteriology, 1995. **177**(4): p. 1094-7.

5. Nakayama, S. and H. Watanabe, *Involvement of CpxA, a sensor of a 2-component regulatory system, in the pH-dependent regulation of expression of Shigella sonnei virF gene.* Journal of Bacteriology, 1995. **177**(17): p. 5062-5069.

6. Durand, J.M.B., et al., *Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of Shigella flexneri: the expression of the virF gene.* Molecular Microbiology, 2000. **35**(4): p. 924-935.

7. Hurt, J.K., S. Olgen, and G.A. Garcia, *Site-specific modification of Shigella flexneri virF mRNA by tRNA-guanine transglycosylase in vitro.* Nucleic Acids Research, 2007. **35**(14): p. 4905-13.

8. Curnow, A.W. and G.A. Garcia, *tRNA-Guanine Transglycosylase from Escherichia coli - Minimal tRNA Structure and Sequence Requirements for Recognition.* Journal of Biological Chemistry, 1995. **270**(29): p. 17264-17267.

9. Eustice, D.C., et al., *A Sensitive Method for the Detection of Beta-Galactosidase in Transfected Mammalian-Cells.* Biotechniques, 1991. **11**(6): p. 739-&.