Genetic Analysis of Allosteric Signaling in RhaR from *Escherichia coli* and Characterization of the VirF Protein from *Shigella flexneri*

By

Bria Collette Kettle

Submitted to the graduate degree program in Molecular Biosciences and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chairperson Dr. Susan M. Egan

_____

Dr. P. Scott Hefty

_____

Dr. Steve Benedict

_____

Dr. David Davido

_____

Dr. Liskin Swint-Kruse

_____

Dr. Edith L. Taylor

Date Defended: November 14, 2013

The Dissertation Committee for Bria Collette Kettle

certifies that this is the approved version of the following dissertation:

Genetic Analysis of Allosteric Signaling in RhaR from *Escherichia coli* and Characterization of the VirF Protein from *Shigella flexneri*

_____

Chairperson Dr. Susan M. Egan

Date approved: December 20, 2013

Abstract

The RhaR and VirF proteins are both members of the large AraC family of bacterial transcription regulators. RhaR activates expression of the *Escherichia coli rhaSR* operon in response to the effector L-rhamnose while VirF is the master regulator of expression of the *Shigella flexneri* type three secretion system (T3SS) and activates transcription in response to temperature and pH. Both proteins consist of two domains: an N-terminal domain (NTD) and a conserved DNA binding domain (DBD) responsible for binding to DNA and contacting RNA polymerase (RNAP) to activate transcription. The RhaR NTD is responsible for protein dimerization and binding the effector L-rhamnose and is required for maximal transcription activation. The VirF NTD is currently uncharacterized, but has been hypothesized to be involved in oligomerization (likely dimerization) of the protein. The principal goals of this study were to further define the mechanism of RhaR interdomain allosteric signaling and to characterize the general mechanisms of transcription activation by VirF.

In the current study, we sought to further elucidate the mechanism of allosteric signaling in RhaR that mediates the response to L-rhamnose. My approach was to examine the role of residues predicted to make interdomain contacts between the RhaR N-terminal domain (NTD) and DNA binding domain (DBD). I generated mutations to examine the role of residues in two regions of the DBD: Allosteric site in subdomain 2 (AS2) and the C-termini of the two helix-turn-helix motifs (C-HTH1 and C-HTH2). At AS2, results indicated that one residue may be involved in inhibitory contacts that reduce the activity of RhaR (-)rhamnose. Furthermore this residue likely interacts with a residue in the RhaR Arm to inhibit transcription activation minus rhamnose. This conclusion is supported by

the isolation of a second-site suppressor mutation in AS2 that restored activity of an Arm variant to wild-type levels. We propose that in the absence of L-rhamnose contacts between the RhaR Arm (in the NTD) and AS2 (in the DBD) regions minus rhamnose that constrain the RhaR conformation such that it is unable to efficiently activate transcription.

We also sought to better characterize the *Shigella* master virulence regulator, VirF. I first investigated the ability of the isolated VirF DBD to activate transcription. The isolated VirF DBD did not activate transcription above background levels, indicating the DBD is not sufficient to activate transcription in the absence of the NTD. We then investigated the role of the VirF NTD. Structural modeling of the VirF protein showed that the NTD of VirF may have structural similarity to the NTDs of the AraC and ToxT proteins, both of which are responsible for dimerization and effector binding. I subsequently screened for potential effectors of VirF and further investigated the oligomeric state of VirF. Preliminary results indicated that VirF likely forms dimers in solution in addition to binding to DNA as a dimer. Furthermore, the effector screen identified bicarbonate as a potential repressor of VirF activity, although more studies are necessary to confirm the role of bicarbonate in VirF activation. Nonetheless, I propose a model where bicarbonate may serve as a spatial regulator of expression of the *Shigella* T3SS, aiding in navigation of the organism to the large intestine where the organism invades the epithelial cells, establishing infection.

The last goal of this study was to investigate inhibition of RhaR and VirF by the small-molecule inhibitor SE-1 *in vitro*. SE-1 inhibits transcription activation *in vivo* and DNA binding *in vitro* of a closely related AraC family regulator, RhaS. I performed electrophoretic mobility shift assays in the presence or absence of SE-1 to determine the ability of the inhibitor to block DNA binding by either RhaR or VirF. I found that SE-1 was

able to inhibit *in vitro* DNA binding by RhaR in a dose-dependent manner. Preliminary studies indicated that SE-1 also inhibited VirF in a dose-dependent manner. From our collective results, we propose that SE-1 blocks transcription activation of RhaS, RhaR and VirF by binding to the conserved DBD and blocking DNA binding. Binding of SE-1 to the conserved DBD that defines the AraC family of activators supports the hypothesis that SE-1 may inhibit other AraC family regulators, providing potential for development as a novel broad-spectrum anti-infective.

Acknowledgements

I would like to begin by thanking my advisor, Dr. Susan M. Egan, for her guidance and support over the past five and a half years, I wouldn't have been able to accomplish this goal without her. I also want to thank the members of my defense committee, Dr. Scott Hefty, Dr. Liskin Swint-Kruse, Dr. David Davido, and Dr. Steve Benedict, for their time, suggestions, guidance and input on my research over the years. I owe a special thanks to Dr. Edith Taylor for filling in as an outside member of my committee at the last moment.

I am especially grateful for my fellow lab members (past and present) for their help, guidance, and most importantly their friendship. I have been incredibly fortunate in that I have gained so many wonderful friends throughout my graduate career. I would not have made it through graduate school without the tremendous amount of support and encouragement I received from so many incredible friends. My family (especially my husband, Mark, and my mom) has been the best support group anyone could ask for. They were so understanding and encouraging at times when many would not have been. I would not be the person I am today, nor would I have accomplished so much without the unconditional love and support that has been provided by my family. Lastly, I want to thank my Grandpa Wilkins for believing in me before I knew how to believe in myself and daring me to dream big.

Table of Contents

Chapter I: Introduction

Transcription regulation is essential in bacterial gene regulation for maximizing utilization of cellular energy. Transcription is carried out by the RNA polymerase (RNAP), which transcribes DNA into messenger RNA (mRNA) and is typically followed by translation of the mRNA into protein. Regulation of gene expression can occur at any stage of gene expression and for a variety of reasons. Regulation of gene expression can aid an organisms response to environmental conditions to optimize energy consumption, prevent interference in cellular processes, or regulate growth and replication (1).

Bacterial Transcription. Transcription occurs in three steps: initiation, elongation, and termination (reviewed in (1), which are performed by the multi-subunit DNA-dependent RNA polymerase (RNAP) in bacteria (2-4). The bacterial core RNAP consists of five subunits (α_2, β, β' and ω) and is capable of transcription, but cannot direct specific transcription initiation through promoter-specific DNA binding (2). Only when the holoenzyme (RNAP core with the σ subunit bound) is formed, can transcription initiation occur, as the σ subunit is responsible for promoter recognition (1, 5-9). In *Escherichia coli*, several different σ factors can initiate transcription (5-8); however σ^{70} is considered the housekeeping σ factor, as it is required for transcription of genes involved in cell growth and basic cell functions (10).

Transcription initiation begins with recognition of the promoter DNA by the RNAP holoenzyme; the σ factor is responsible for promoter recognition based on the consensus sequences present in the promoter sequence (1, 9). The σ^{70} subunit recognizes two regions in the promoter DNA, the -35 and the -10 sequences, which are centered at 35 and 10 base pairs upstream of the transcription start site, respectively. The σ^{70} -35 consensus sequence

is 5'-TTGACA-3' and the -10 consensus sequence is 5'-TATAAT-3' (9, 11). Isomerization occurs following promoter binding by the RNAP holoenzyme (2). During isomerization, open complex formation occurs through unwinding of the duplexed DNA strands near the transcription start site (2, 12-14). Proceeding open complex formation, the RNAP holoenzyme begins the process of transcribing the DNA in the elongation step, ending with termination of transcription.

Transcription Activation. Promoter recognition and binding by the RNAP holoenzyme does not occur equally for all promoters (2). Variations in interactions of various promoter elements with the RNAP holoenzyme at the step of promoter recognition can result in weaker affinity of RNAP for a given promoter, however differences in affinity do not allow the bacterium to regulate expression in response to environmental conditions (2, 15). Therefore, many promoters have specific activator proteins that bind in addition to RNAP holoenzyme to increase the rate of transcription initiation, allowing adaptive regulation (2, 15-17). Transcription activation can involve a single activator protein or multiple activators that can either interact with the promoter DNA to improve interaction with RNAP, or the activator can interact with RNAP to alter promoter preference (2). At single activator-dependent promoters, activators typically bind upstream of or slightly overlapping the -35 and -10 elements directly interacting with RNAP to facilitate transcription initiation (2). Interaction of the transcription activators with RNAP can occur in various manners. Activators can bind upstream of the promoter and interact with the αCTD of RNAP (2, 16). Transcription activators can also overlap the -35 element when bound at promoter DNA, interacting with either domain 4 of the σ subunit of RNAP (2, 16, 18), the αCTD upstream of the activator (16, 17) or with the αNTD subunit (19).

Furthermore, other activators can alter the promoter conformation such that RNAP is better able to interact with the -35 and -10 elements (20, 21). Activation involving multiple activators is more complex in that the activators typically bind independently (2). Regulation by multiple activators can involve displacement of a repressor by one of the components, allowing the other activator to contact RNAP to initiate transcription (2). Alternatively, a secondary activator can be required to position the primary activator such that it properly contacts RNAP to activate transcription (2).

Transcription Activation by Cyclic AMP Receptor Protein (CRP). More than one hundred promoters in *E. coli* are regulated by cyclic AMP receptor protein (CRP) (also known as catabolite activator protein or CAP), a global transcription activator in *E. coli* (22, 23). CRP activates transcription in the presence of its effector cyclic AMP (cAMP) by binding to specific sites on DNA in or near target promoters, enhancing the ability of RNAP holoenzyme to activate transcription (22, 23). With cAMP bound, the protein binds to DNA as a dimer, bending the DNA by ~80° (22). There are three classes of CRP-dependent promoters, which are determined based on the position of CRP binding at the promoter (23). Both Class I and Class II CRP-dependent promoters require only CRP for activation and have a single CRP binding site. At Class I CRP-dependent promoters, the CRP binding site is upstream of the RNAP binding site and on the same face of the DNA. At Class II CRP-dependent promoters, the CRP binding site overlaps the RNAP binding site, apparently replacing the -35 promoter element. Class III CRP-dependent promoters require multiple activator molecules for full transcription activation (i.e. two or more CRP molecules or one or more CRP molecules in addition to one or more regulon-specific activator proteins) (23).

AraC Family of Transcription Activators. The AraC family is a large family of bacterial transcription regulators that have been identified in seventy percent of sequenced bacterial genomes, including a wide diversity of bacterial organisms (24). Members of the AraC family are defined by a DNA binding domain (DBD) that consists of a conserved region of ~100 amino acids (25-29). The conserved DBD is responsible not only for binding to target DNA, but also for transcription activation through contacts with the α subunit of RNAP (30-36) and/or residues at the C-terminal end of σ^{70} (37-41). In addition to sequence conservation within the DBD, the tertiary structure of this domain is conserved. The DBD consists of two helix-turn-helix (HTH) motifs each made up of three α-helices connected by a longer central α-helix (42). Each of the HTH motifs contacts an adjacent major groove in the DNA, both on the same face of the DNA, as seen in the MarA co-crystal with DNA (43). While some members of the AraC family consist only of the DBD, such as MarA and SoxS (29), a majority of AraC family members also possess a second domain. A large subset of homologs share sequence similarity with the effector binding and dimerization domain of AraC - the founding member of the family (28, 29). This subset includes the regulators of the *Escherichia coli* L-rhamnose regulon, RhaS and RhaR, and VirF, the master regulator of the type three secretion system (T3SS) in *Shigella flexneri*.

AraC family proteins most commonly control the expression of genes involved in processes such as carbon metabolism, stress responses, and virulence (24, 26, 29, 42). Like AraC, RhaS and RhaR regulate expression of genes whose products are involved in carbon metabolism. Other examples of AraC family members in this class (and the function of the genes they regulate) include XylS from the TOL plasmid of *Pseudomonas putida* (metabolism of benzene derivatives) (44), MelR (metabolism of melibiose) (45) and XylR

(metabolism of xylose) (29) from *E. coli*. The second class of AraC family regulators includes those that regulate stress responses, such as Ada (response to alkylating agents) from *E. coli* (46), *Salmonella typhimurium* (47) and *Bacillus subtilis* (48), and SoxS (response to oxidative stress) (49) and MarA (response to antibiotics) (50) from *E. coli*. The third class of AraC family regulators includes those that regulate expression of genes associated with virulence. Examples of AraC family virulence regulators include ToxT of *Vibrio cholerae* (cholera toxin, the toxin-coregulated pilus, and it's own expression) (51-55), Rns from Enterotoxigenic *E. coli* (ETEC) (CS1 and CS2 pili) (56), and VirF from *Shigella flexneri* (master regulator of the type three secretion system (T3SS)) (57).

Many AraC family proteins are utilized by pathogenic bacteria to regulate expression of virulence genes and are often required for expression of multiple virulence factors (29). Bacterial virulence is reduced without affecting bacterial growth in strains where mutations disrupting the function of the AraC family virulence regulator are made (58, 59). Thus, we hypothesize that AraC family virulence regulators may be an advantageous target for novel antimicrobials, as targeting these proteins would place less selective pressure on the organism to develop resistance due to their non-essential nature (60-63).

Effector Binding and Responses by AraC Family Members. Many proteins within the AraC family respond to environmental small molecule effectors that regulate their activity. The most well characterized member of this class is the AraC protein itself. AraC responds to its effector arabinose to regulate expression of the *araBAD* operon using a well-characterized molecular mechanism known as the "light switch" mechanism (Figure 1) (64-68). In addition, high-resolution structures are available for both domains of

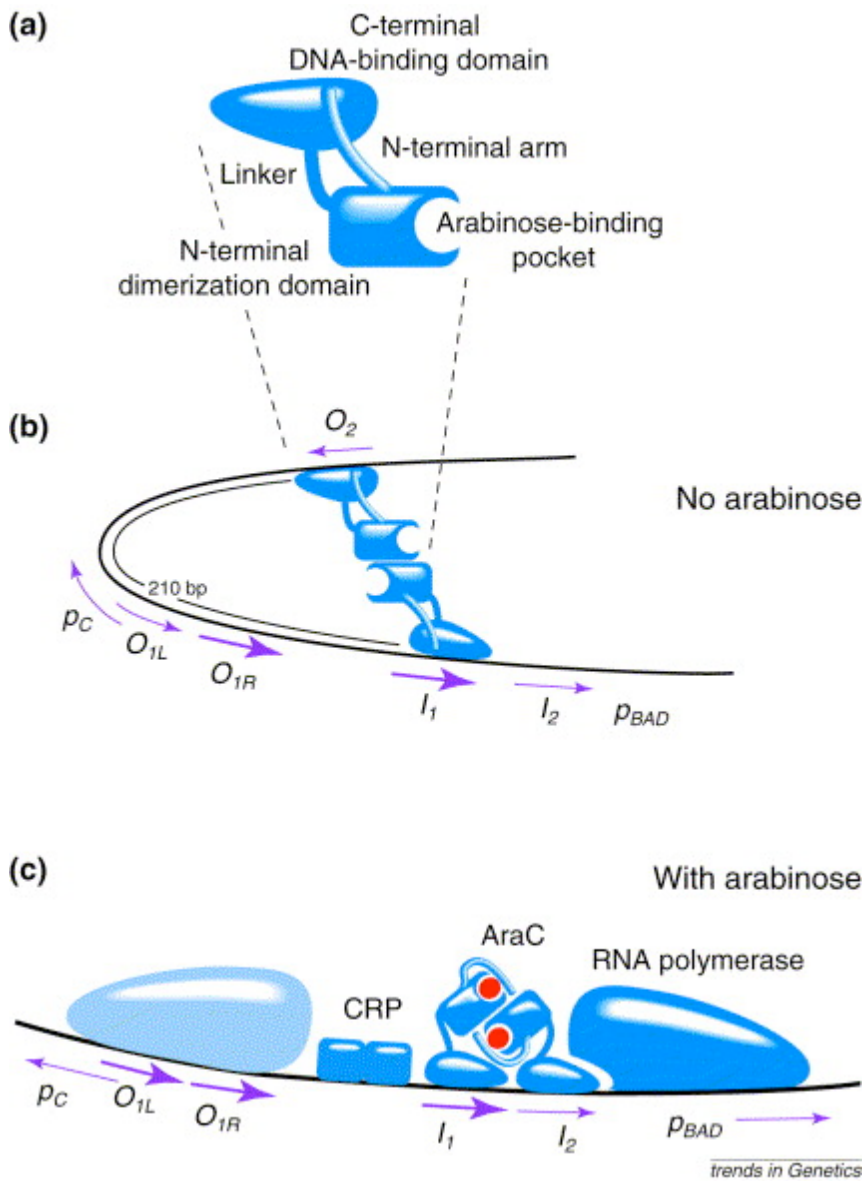


Figure 1. The AraC "Light switch" Mechanism. A. In the absence of the effector arabinose, the N-terminal Arm of AraC contacts the DNA binding Domain (DBD), constraining the conformation of the protein such that one monomer contacts the distal O_2 half-site, resulting in DNA looping and inhibition of *araBAD* transcription. **B.** In the presence of arabinose, the Arm folds over the arabinose-binding pocket, allowing the monomers to contact the I_1 and I_2 half-sites, thus activating transcription. From (69)

AraC, including structures of the NTD in the absence and presence of the effector arabinose, although no full-length structure is yet available (70-72). In the presence of arabinose, the AraC Arm residues bind over the arabinose-binding pocket in the NTD (71). In this state, the two domains of AraC are predicted to adopt a conformation in which the two DBDs of an AraC dimer simultaneously bind to the direct repeat $araI_1$ and $araI_2$ half-sites at *araBAD*, thus activating transcription (66, 67, 73, 74). In the absence of arabinose, the AraC Arm was unstructured in crystals of the AraC NTD (71). When arabinose is not bound, the AraC Arm is predicted to contact the DBD, constraining the protein conformation such that one monomer contacts a distal half-site $araO_2$ at 211 base pairs upstream of $araI_1$, resulting in DNA looping and inhibition of *araBAD* transcription (64, 67, 68, 73, 75-78). Much evidence for the role of the AraC Arm and its contact site in the DBD comes from deletions in the Arm and point mutations in each region that confer constitutive activation (64, 68, 71).

ToxT is another example of an AraC family protein that responds to an environmental effector. Unlike AraC, ToxT is in its non-activating state in the presence of its effector (79, 80). The crystal structure of the full-length ToxT protein in its non-activating state with effector bound has been solved (79). The crystal structure of ToxT includes a bound fatty acid effector, a very large (~2000 Å^2) polar interdomain interface, and is described as a "closed" conformation (79). In this non-activating state, NTD residues that surround the entrance to the ToxT effector-binding pocket, as well as the fatty acid effector itself, are positioned to contact the DBD (79). Effector-dependent changes in the interdomain contacts between residues in the ToxT NTD and DBD provide a likely mechanism for transmission of the effector-binding status in ToxT. Indeed, it has been hypothesized that decoupling of interdomain interactions in the absence of effector results

in formation of an "open" ToxT conformation that is capable of binding DNA and activating transcription (79). Alternatively, given the very large ToxT interdomain interface (79) and the recent finding that the two domains of AraC make contacts both in the activating and non-activating states (74), it is possible that the ToxT interdomain contacts are altered/rearranged, but not eliminated, in its activating state.

Both sequence and structural alignments of AraC with ToxT indicate that the AraC DBD residues that are contacted by the Arm (NTD) in its non-activating state overlap with ToxT DBD residues that contact the ToxT effector and residues around the effector-binding pocket (NTD) (Figure 2). This suggested that the (-)arabinose state of AraC may be similar, both structurally and functionally, to the "closed", effector-bound state of ToxT (79). We refer to these regions of the DBDs of AraC and ToxT as the allosteric site in subdomain 2 (AS2). This similarity between the AraC and ToxT mechanism further supports the hypothesis that this "closed", non-activating state might be a feature in common with additional AraC family proteins. Evidence suggests that XylS also utilizes a "closed", non-activating state in the absence of its effector, although a specific mechanism for inhibition has not been proposed (81-84). Although there is no evidence for DNA looping by ToxT or XylS, a protein conformation similar to the proposed AraC (-)arabinose state might prevent binding to the adjacent DNA half-sites required to activate transcription.

The L-Rhamnose Regulon. The L-rhamnose regulon of *E. coli* consists of three operons: *rhaSR*, *rhaBAD* and *rhaT* (Figure 3) (25, 85-88). The *rhaSR* operon encodes the RhaS and RhaR proteins, which are both AraC homologs (25-27, 29). The structural genes required for L-rhamnose catabolism are encoded in the *rhaBAD* operon and encode the

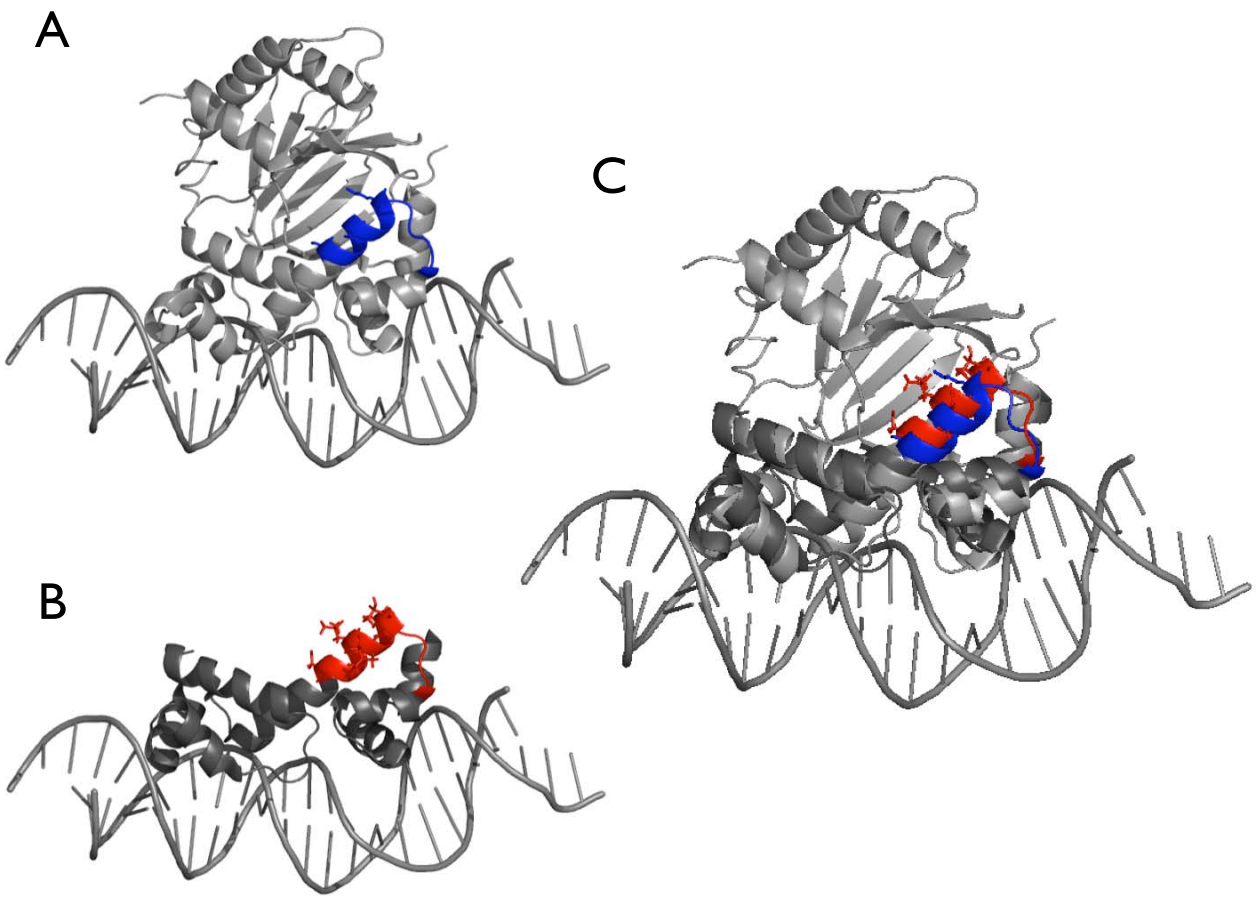


Figure 2. Allosteric Site in Subdomain 2 (AS2) of the DBDs of AraC and ToxT.

A. Full-length ToxT crystal structure (79) with residues in AS2 highlighted in blue and modeled on DNA from the MarA co-crystal structure. Side chains of residues involved in interdomain contacts and/or effector contacts are shown as sticks. **B.** AraC DBD structure (72) with residues in AS2 highlighted in red and modeled on DNA from the MarA co-crystal structure. Side chains of residues predicted to interact with the AraC Arm shown as sticks. **C.** Structural alignment of the AraC DBD (dark gray) with the ToxT crystal structure (light gray). Proteins were aligned and modeled on DNA from the MarA co-crystal structure using the MacPyMOL Molecular Graphics System.

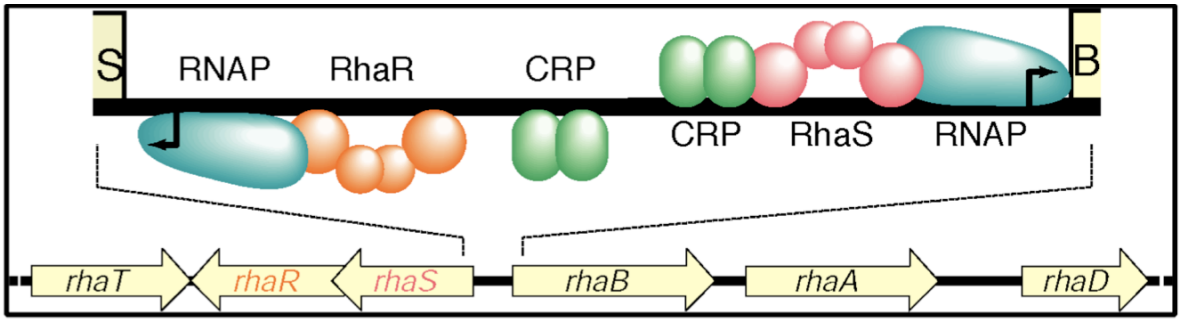


Figure 3. Schematic Representation of the *Escherichia coli* L-rhamnose Regulon.

Thick horizontal black lines represent DNA. Top: Expanded view of the regulatory region between the *rhaSR* and *rhaBAD* operons with the approximate positions of the known proteins bound to DNA at *rhaBAD* and *rhaSR*. Bent arrows indicate transcription start sites. RhaR is in orange, RhaS in pink, CRP in green, and RNAP in teal. Letters "S" and "B" at the ends of the top line represent the beginning of the *rhaS* and *rhaB* genes, respectively. Bottom: Representation of the orientation and direction of transcription for the genes in the L-rhamnose regulon. Courtesy of Susan Egan.

L-rhamulose kinase, L-rhamnose isomerase, and L-rhamulose 1-phosphate aldolase, respectively (87, 89). L-rhamnose is transported into the cell by the L-rhamnose-proton (H^+) symporter encoded by the *rhaT* operon (88). Expression of both the *rhaBAD* and the *rhaT* operons is activated by RhaS in response to L-rhamnose (86, 88). RhaR activates RhaS and RhaR. RhaS and RhaR activate transcription to maximal levels (above basal levels) only in the presence of their common effector, the sugar L-rhamnose. In the absence of L-rhamnose, all three of the L-rhamnose-regulated operons in *E. coli* are expressed at low, basal levels (86, 90, 91). When L-rhamnose becomes available, basal levels of RhaR protein activate expression of the *rhaSR* operon (25, 86). The increased level of RhaS protein then activates transcription of the *rhaBAD* and *rhaT* operons (86-88, 92). All three operons within the L-rhamnose regulon require CRP in addition to RhaS or RhaR for maximal activation (31, 86, 91). CRP binds upstream of RhaS and RhaR at a position centered at -92.5 and -93.5 at *rhaBAD* and *rhaT* (86, 91), respectively, and at -111.5 at *rhaSR* (86).

RhaS and RhaR: Activators of the L-Rhamnose Regulon. RhaS and RhaR activate transcription to maximal levels (above basal levels) only in the presence of their common effector, the sugar L-rhamnose. Like AraC, the RhaS and RhaR NTDs are responsible for binding L-rhamnose and dimerization (93) and [A, Kolin, G.K. Hunjan and S.M. Egan unpublished results]. The DBDs of RhaS and RhaR bind DNA and contact RNAP to activate transcription (25, 37, 94, 95). RhaS and RhaR dimers bind to their respective promoters at two 17 base pair half-sites separated by 16 or 17 base pairs, respectively, that overlap the -35 promoter element by four base pairs (86, 94). In the absence of L-rhamnose, full-length RhaS is unable to bind DNA; however the isolated RhaS DBD can bind DNA and activate transcription (86, 93). This suggests that the NTD in the absence of L-rhamnose

limits RhaS binding to DNA; however in the presence of L-rhamnose, inhibition of DNA binding is somehow relieved. In contrast, both RhaR and the isolated RhaR DBD can bind DNA in the absence of L-rhamnose. Nevertheless, only full-length RhaR in the presence of L-rhamnose significantly activates transcription, suggesting that in the absence of L-rhamnose RhaR is limited in its ability to contact RNAP (90, 93, 94). Both RhaS and RhaR make protein-protein interactions with the σ^{70} subunit of RNAP (37, 38).

The current evidence supports a model in which L-rhamnose binding to RhaR increases both DNA binding (although with a relatively modest increase in binding affinity) and transcription activation (90, 93, 94), with the latter involving contacts with RNAP σ^{70} and perhaps also the C-terminal domain of the RNAP α-subunit (37, 95). These L-rhamnose-dependent increases in DNA binding and transcription activation are likely due to structural changes in the RhaR DBD, whereas L-rhamnose binds to the RhaR NTD. Therefore, an allosteric signal must be transmitted between the RhaR domains to convert RhaR between its (-)rhamnose and (+)rhamnose states. Our previous work suggests that the linker that connects the two RhaR domains does not play a direct role in transcription activation (96). Thus, the allosteric signal that converts RhaR from its (-)rhamnose state to its (+)rhamnose state likely involves L-rhamnose-dependent changes in the interdomain contacts.

VirF, Master Regulator of the *S. flexneri* T3SS. The AraC family activator VirF is encoded on the large virulence plasmid of the human pathogen *S. flexneri*, outside of the entry region (97). Expression of VirF begins a regulatory cascade, resulting in expression of several operons encoding products required for virulence functions such as invasion and intercellular spread of the organism (98-102). Expression of the *virF* gene is temperature

dependent, with transcription at 37°C, but not at 30°C, which is mediated by H-NS repression (103-105): at temperatures below 32°C, the *virF* promoter is intrinsically curved, allowing the nucleoid associated protein H-NS to interact with two sites at the *virF* promoter, blocking transcription (103). At temperatures above 32°C, the structure of the DNA changes, releasing H-NS and allowing transcription of *virF* (103, 105). At permissive temperatures, VirF directly activates expression of the *virB* (106, 107) and *icsA* genes (107, 108). The *virB* gene product, VirB, is also a transcriptional regulator, which acts as an anti-repressor to alleviate H-NS-mediated inhibition of the operons encoding the type three secretion system (T3SS) and many of its effectors (including the *mxi*, *spa*, and *ipa* operons) (109, 110).

***Shigella flexneri* Pathogenesis.** *Shigella flexneri* is the etiological agent of bacillary dysentery, or shigellosis, in humans. Symptoms of shigellosis include watery diarrhea, fever, cramping, with blood and mucous present in the stool later in infection. Worldwide, *Shigella* is the cause of more than 120 million illnesses annually, mostly in developing nations, and approximately 1.1 million deaths, mostly of children under the age of 5 (111). In the U.S., *Shigella* is responsible for approximately 500,000 infections annually (112). Of the annual *Shigella* infections in the US, 27,000 are classified as drug resistant, which constitutes a serious antibiotic resistance threat (112). Of the *Shigella* species, *S. flexneri* causes the highest mortality and is endemic in most developing countries (113). *S. flexneri* is transmitted human-to-human via the fecal-oral route; fewer than a hundred viable organisms are capable of establishing an infection (114, 115). After ingestion of

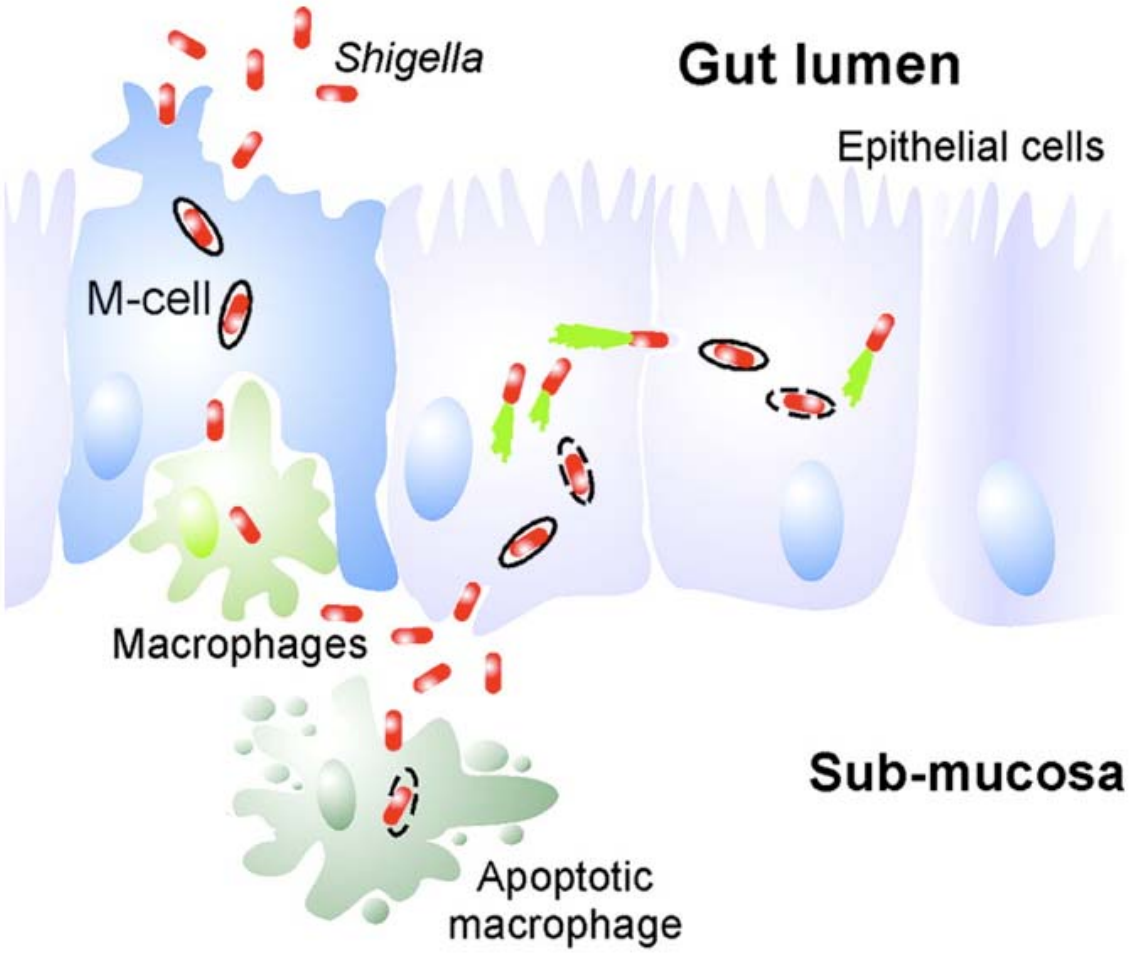


Figure 4. Cellular Pathogenesis of *Shigella flexneri*. *S. flexneri* crosses the epithelial layer by transcytosis through M cells, where it is phagocytosed by macrophages. *S. flexneri* escapes the macrophage by inducing apoptosis and invades the epithelial cells at the basolateral pole. *S. flexneri* then escapes the phagosome into the epithelial cell cytoplasm where it replicates. *S. flexneri* can then subvert the host actin cytoskeleton for intracellular and intercellular motility. Adapted from (116).

contaminated food or water, organisms pass through the stomach and small intestine to the colon, where infection is established (116).

In the colon, *S. flexneri* establishes infection by promoting its uptake into the microfold cells (M cells) and is transcytosed across the epithelial layer to the basolateral pole of the intestinal epithelia (Figure 4) (117-119). M cells specialize in sampling particles from the intestinal tract and delivering them to underlying mucosal lymphoid tissue, where immune responses are mediated (120). Once across the epithelial layer, *S. flexneri* are endocytosed by macrophages where the organisms induce apoptosis (121-123). After escaping the macrophage, *S. flexneri* invades the epithelial cells from the basolateral side by inducing endocytosis, they then escape the phagolysosome, and replicate within the host cytoplasm (124). *S. flexneri* can also subvert host cell actin polymerization for intracellular motility and cell-to-cell spread (125-128).

Factors required for *S. flexneri* virulence are encoded in a 31kb region of the large virulence plasmid (99, 129, 130), termed the entry region, which encodes the type three secretion system (T3SS) in addition to other virulence factors (130, 131). Expression of the genes required for invasion is temperature dependent, as bacteria that are phenotypically invasive at 37°C are non-invasive when grown at 30°C. Temperature regulation is physiologically relevant, as increased temperature signifies entry of the bacterium into the host (130).

Type Three Secretion System of *Shigella flexneri*. The type three secretion system (T3SS) of *S. flexneri* is responsible for delivery of bacterial protein effectors into the host cytosol (Figure 5) (132). The *Shigella* T3SS is required for invasion of the host colonic epithelial cells and cell-to-cell spread, as well as manipulation of the host innate and

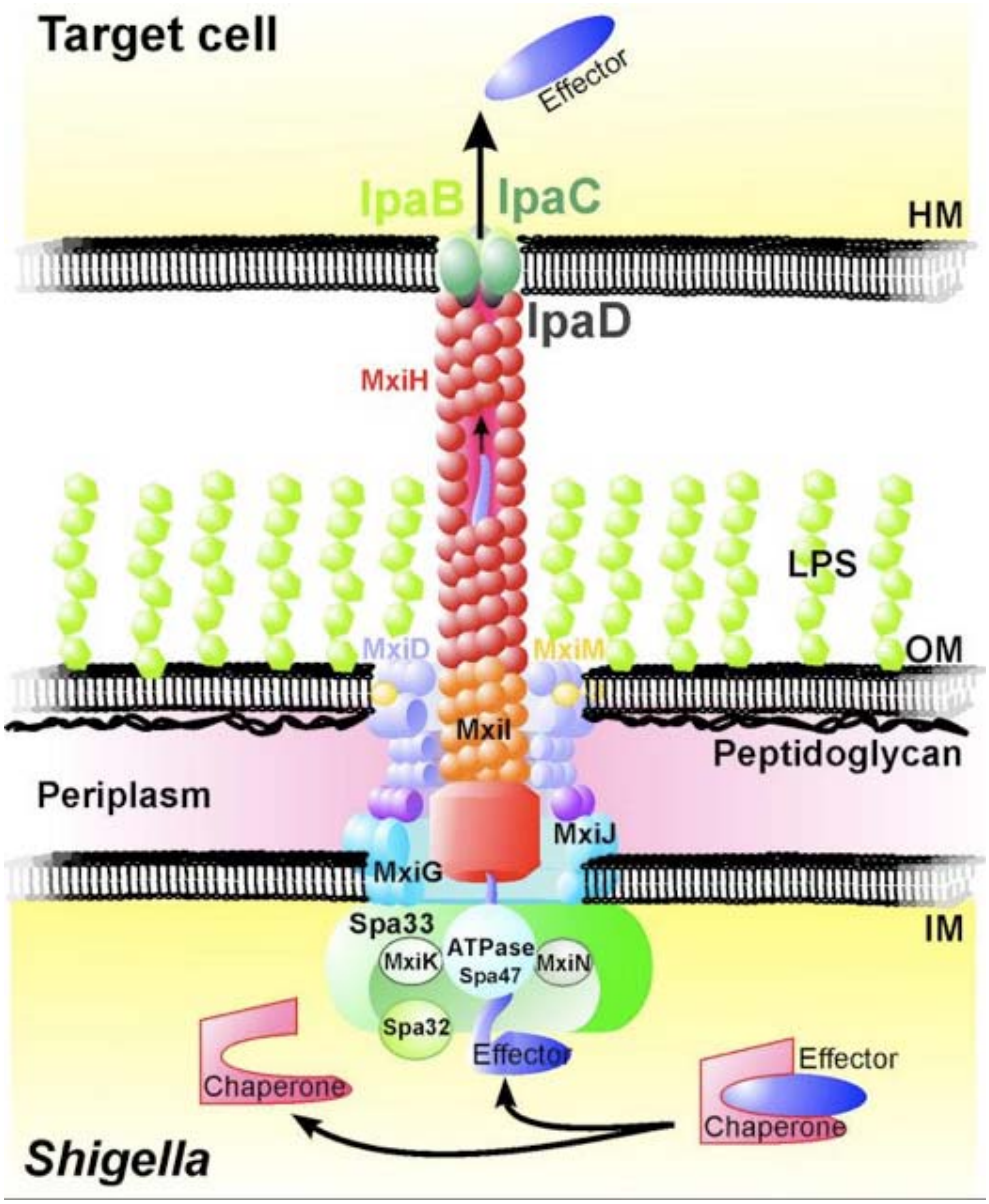


Figure 5. Architecture of the *S. flexneri* T3SS. The *S. flexneri* T3SS consists of a seven-ringed basal body spanning the bacterial inner membrane (IM), periplasm, peptidoglycan and outer membrane (OM). The hollow needle is attached to a socket and protrudes from the basal body to the bacterial surface. Contact with host cell membranes (HM) triggers the IpaD-guided membrane insertion of the IpaB-IpaC translocon at the needle tip. From (116).

adaptive immune responses (133-135). The *Shigella* T3SS is comprised of approximately 22 proteins arranged into a needle-like protein oligomer anchored in the bacterial membrane (135). The needle-like structure delivers T3S effectors across three membranes: both the inner and outer membrane of the Gram-negative bacterium and across the host cytoplasmic membrane into the host cytosol (116, 135-137).

Assembly of the T3SS needle complex occurs in a distinct order. First, the seven-ringed basal body is assembled, spanning the bacterial inner membrane, periplasm, peptidoglycan layer and outer membrane (116). After assembly of the basal body, the needle is then assembled. The *Shigella* T3SS needle is composed of polymers of the MxiH protein assembled in a helical polymer that extends into extracellular space. The needle is assembled as MxiH is secreted from the basal body, a process that is regulated by a number of cytoplasmic proteins associated with the inner membrane ring of the basal body (116). Secretion of the needle tip complex (composed of IpaD, IpaB, and IpaC) is the next step in formation of the T3SS of *Shigella*. The tip complex is responsible for controlled secretion and translocation of effector proteins into the host cell (138, 139). In the absence of a host cell, IpaD localizes to the needle tip and interacts with IpaB inside the needle channel to block secretion (139-142). Upon host cell contact, IpaD no longer blocks secretion, allowing both IpaB, and subsequently IpaC, to be secreted. With activation of the T3SS, IpaB inserts into the host membrane along with IpaC to form the translocation pore into the host cell (132, 139, 143). Once the translocation pore has been assembled in the host membrane, the needle is in an "open" conformation allowing secretion of effector proteins into the host (116).

Research Goals. The overall goal of this study was to further define the mechanism of transcription activation of RhaR and to better characterize the AraC family virulence regulator VirF. While RhaR is well studied and well characterized, the allosteric mechanism that regulates transcription activation in response to L-rhamnose has not been delineated. Therefore, the first goal of this study was to identify residues that may be involved in the RhaR allosteric response to L-rhamnose. To identify these residues, I made mutations at residues predicted to be located at the interdomain interface of the RhaR DBD (based on alignments with AraC and ToxT and upon the crystal structure of ToxT (79)). Variants were tested *in vivo* for activation defects either in the presence or absence of L-rhamnose. From the mutagenesis, I identified one residue, H269, which likely makes inhibitory contacts in the absence of the effector L-rhamnose. I also found evidence for interaction of this residue with residue L35 in the RhaR Arm with isolation of a second-site suppressor, L35K-H269V.

The second goal of this study was to determine the role of the N-terminal Extension of RhaR. I first made N-terminal truncations of the RhaR protein similar to those previously described due to concern that increased protein levels may mask potential defects of the truncated variants (144). I found that the Extension is not required for stimulation (+)rhamnose; however, it may have a role in maintaining protein stability and/or function (-)rhamnose.

The third goal of this study was to further characterize the AraC family virulence regulator VirF. I first evaluated the requirement for the N-terminal domain of VirF for transcription activation and determined that the VirF DBD is not sufficient to activate transcription *in vivo*. I next screened for potential chemical effectors of VirF as it had

recently been found that two closely related AraC family virulence regulators (ToxT and RegA) respond to chemical effectors (79, 80, 145, 146). Bicarbonate decreased VirF activation at physiological levels, suggesting this compound may be a physiological repressor of VirF. I next examined the ability of VirF to oligomerize *in vitro*. This first required purification of the VirF protein. An MBP-VirF fusion was generated and the purification method optimized, allowing me to perform *in vitro* tests of DNA binding and analytical size-exclusion chromatography of the protein in solution. These studies indicated that VirF likely binds DNA as a dimer and forms a dimer in solution.

Lastly, I also examined the ability of a small molecule inhibitor, SE-1, to inhibit DNA binding by $GB1^{b}$-RhaR and $MBP\text{-}VirF^{co}$ *in vitro*. SE-1 has been shown to inhibit transcription activation by RhaR and VirF *in vivo* (147, 148), therefore I tested the ability of the inhibitor to block DNA binding to aid in determination of the mechanism of SE-1 inhibition of transcription activation. SE-1 was able to block DNA binding by both RhaR and VirF *in vitro* in a dose-dependent manner (147, 148). Additionally, our findings indicate that SE-1 is likely specific to the conserved DBD that defines the AraC family of activators and may have potential for development into a novel anti-microbial therapy in the future.

Chapter II: Materials and Methods

Culture Media and Conditions. *Escherichia* cultures for β-galactosidase assays were grown in MOPS [3-(*N*-morpholino)propanesulphonic acid]-buffered minimal medium (149, 150) with a carbon source of 0.4% glycerol and in the presence or absence of 0.4% L-rhamnose (38). Most other cultures were grown in Tryptone-yeast extract medium (0.8% tryptone, 0.8% yeast extract, 0.5% NaCl, pH7.0). Tryptone Broth (0.8% tryptone, 0.05% NaCl, pH 7.0) with 5mM $CaCl_2$ was used to grow cultures in preparation for phage P1 infection or transduction. For solid media, cells were grown on Nutrient agar plates (2.3%Difco Nutrient agar, 0.5% NaCl) with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside, 40mg/ml) and with or without 0.4% L-rhamnose to test *lacZ* expression. *Shigella flexneri* cultures were grown in Luria Bertani (LB) broth (1.0% tryptone, 0.5% yeast extract, 1% NaCl, pH 7.0), MOPS-buffered medium, or on Tryptic soy agar (TSA) (Becton, Dickinson and Company, BD, Cockeysville, MD) containing Congo red dye (0.025%). All *S. flexneri* strains were picked as red colonies from TSA containing Congo red. Ampicillin (200 μg/ml), kanamycin (50μg/ml), chloramphenicol (50μg/ml) and tetracycline (20μg/ml) were added as necessary. All cultures were grown at 37°C with aeration.

General Methods. Standard methods for restriction endonuclease digestion, ligation, and transformation were used in this study. Restriction endonucleases and T4 DNA ligase were from New England Biolabs (Beverly, MA). Oligonucleotide primers were synthesized by Eurofins MWG Operon (Huntsville, AL). DNA fragments were amplified for cloning using either the Expand High Fidelity PCR System (Roche Indianapolis, IN) or Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). PCR products were

Table 1. Oligonucleotides used in this study

Oligo Number	Sequence	Use
2082	GTGAAGCTTTTAATCTTTCTGCGAATTGAG	Clone wild type *rhaR* into pHG165 (downstream)
2452	GATGAATTCGGTCACCGCGTGATATTCGCC	Clone wild type *rhaR* into pHG165 (upstream)
2793	CGATGACTCTTCAGGTGTAACGATGGCGATTC	middle of *rhaR* (downstream) Splice at S196
2794	CGATGACTCTTCAACCAGTGATTCGTTGCCG	middle of rhaR (upstream) Splice at S196
2905	CACCTCTTCCTCTGACCTGTCGCAGATATT	Construct *rhaR* H262X (middle upstream)
2907	GCGCTCTTCCATGACACACTCTGACCTGTC	Construct *rhaR* Y265X (middle upstream)
2908	ATCCTCTTCTCTCCAG**NNN**AGCCGCCT	Construct *rhaR* H269X (middle downstream)
2909	GGTCTCTTCGGAGAAGATATTGCGCAT	Construct *rhaR* H269X and S270X (middle upstream)
2914	GTCCTCTTCCAGAGTGTGT**NNN**GCGCAATATCTTCTCCAGCATAG	Construct *rhaR* H262X (middle downstream)
2915	GTGCTCTTCTCATGCGCAA**NNN**CTTCTCCAGCATAGCCG	Construct *rhaR* Y265X (middle downstream)
2916	ATCCTCTTCTCTCCAG**NNN**AGCCGCCTGTTAATCAGTGATAT	Construct *rhaR* H269X (middle downstream)
2917	ATCCTCTTCTCTCCAGCAT**NNN**CGCCTGTTAATCAGTGATAT	Construct *rhaR* S270X (middle downstream)
2918	GTGCTCTTCTCATGCGCAATAT**NNN**CTCCAGCATAGCCG	Construct *rhaR* L266X (middle downstream)
2937	TTACTCTTCACTGATGCGCCACCGTGGC	Construct rhaR L37X-H269Variant (middle upstream)
2938	GCACTCTTCTCAGTTAAAA**NNN**CTCAAAGATGATTTTTTT	Construct rhaR L37X-H269Variant (middle downstream)
2954	GGCCTCTTCTATGCTGGAGAAGATATTGCGCATGACAC	Construct *rhaR* R721X, L272X, or L273X (middle upstream)
2955	CCACTCTTCGCATAGC**NNN**CTGTTAATCAGTGATATTTGCACCG	Construct *rhaR* R271X (middle downstream)
2956	CCACTCTTCGCATAGCCGC**NNN**TTAATCAGTGATATTTGCACCG	Construct *rhaR L*272X (middle downstream)
2957	CCA CTC TTCGCATAGCCGCCTG**NNN**ATCAGTGATATTTGC	Construct *rhaR L*273X (middle downstream)
2958	GATGAATTCTAAATATAGTTTGGTTATTCTGTTGAATTTATG	Clone wild type *virF* into pHG165 (upstream)
2968	GTGGAATTCTATTCTGTTGAATTTATGGATCAGATAAGGAAGATTGTTGAAAAAAACATC	Construct Δ2-160 *virF* (upstream)
*2969	GTGGAATTCGCCAGGATGTTAAAACTTCTCAAAGATGATTTTTTTGCC	Construct Δ34-AS2 variants (upstream)
2980	TCTCTCTTCGCGCATCAG**NNN**AAACTTCTCAAAGATGAT	Construct rhaR L35X-H269Variant (middle downstream)
2982	GAACTCTTCTGCGCCACCGTGGCTACCTCGGCCAGAG	Construct rhaR L35X-H269Variant (middle upstream)

Oligo Number	Sequence	Use
3014	GTGCTCTTCATATCTTCTC**NNN**CATAGCCGCCTGTTAATCAGTGATAT	Construct *rhaR* Q268X (middle downstream)
3017A	GTGGAATTCCGCGTGATATTCGCCAGGGACGGGATGTTAAAACTTCTCAAAGATGATTTTTTGCC	Construct *rhaR* Δ2-34 (upstream)
3018A	GTGGAATTCCGCGTGATATTCGCCAGGGACGGGATGGATTTTTTTGCCAGCGACCAGCAG	Construct *rhaR* Δ2-40 (upstream)
3019A	GATGAATTCCGCGTGATATTCGCCAGGGACGGGATG	Construct *rhaR* wild type with same SD as N-terminal deletions (upstream)
3037	GTGCTCTTCGATATTGCGCATGACACACTCTGAC	Construct *rhaR* Q268X (middle upstream)
3042	GTGCTGCAGTTAAAATTTTTTATGATATAAGTAAAATTTCTTTGGAG	Clone wild type *virF* into pHG165 (downstream)
3043	GTGAAGCTTTTA**ACG**TTTCTGCGAATTGAG	Construct *rhaR* D312R
3045	GTGAAGCTTTTACTGCGAATTGAGATGACG	Construct *rhaR* Δ311-312
3046	GTGAAGCTT*TTAACGCAGCAG*ATCTTTCTGCGAATTGAG	Construct *rhaR* +LLR(stop)
3047	GTGAAGCTT*TTAACGNNNNNNNNN*ATCTTTCTGCGAATTGAG	Construct *rhaR* +NNNR(Stop)
3048	GCGCATATGATGGATATGGGACATAAAAACAAAATAGATATAAAGG	Clone wild type *virF* into pDZ1 and pDZ3 (upstream)
3049	GCGGGATCCTTAAAATTTTTTATGATATAAGTAAAATTTCTTTGGAG	Clone wild type *virF* into pDZ1 (downstream)
3050	GCGGGATCCAAATTTTTTATGATATAAGTAAAATTTCTTTGGAG	Clone wild type *virF* into pDZ3 (downstream)
3051	GTGAAGCTT*TTAACGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN*ATCTTTCTGCGAATTGAGATGACG	Construct *rhaR* + 11(X)R(stop)
3052	GTGAAGCTTTTAATC**TTC**CTGCGAATTGAGATGACG	Construct *rhaR* K311E
3056	[IRD700]CGCTGTATCTTGAAAAATCGACGTTTTTTACGTGGTTTTCCGTCGAAAATTTAAGGTAAGAAC	IRD700 Labeled Oligo encoding full RhaR binding site at rhaSR (top strand)
3057	[IRD700]GTTCTTACCTTAAATTTTCGACGGAAAACCACGTAAAAAACGTCGATTTTTCAAGATACAGCG	IRD700 Labeled Oligo encoding full RhaR binding site at rhaSR (btm strand)
3068	GCGCATATGGATCAGATAAGGAAGATTGTTGAAAAAAACATCGAG	Clone wild type *virF*-DBD into pDZ1 and pDZ3 (upstream)
3073	GCGAAGCTTGCTCTTCACTGATCTTTCTGCGAATTGAGATGACGCCACTG	Construct *rhaR-DBD*+arm (middle upstream)
3074	GCGGAATTCATATTCGCCAGGGACGGGATGAGTGATTCGTTGCCGCC	Construct *rhaR* DBD
3086	GCGGCTCTTC*TCAGTTAAAACTTCTCAAAGATGATTTTTTTTAA*AAGCTTCGC	Construct *rhaR*-DBD+arm (top strand of "arm")
3087	GCGAAGCTT*TTAAAAAAAATCATCTTTGAGAAGTTTTAACTGA*GAAGAGCCGC	Construct *rhaR*-DBD+arm (btm strand of "arm")
3092	GTCCTCTTCCAGAGTGTGTCATGCG**NNN**TATCTTCTCCAGCATAG	Construct rhaR Q264X (middle downstream)
3093	GTGCTCTTCTCATGCGCAATATCTT**NNN**CAGCATAGCCGCCTG	Construct *rhaR* L267X (middle downstream)
3094	GTGAAGCTTTTAATCTTTCTGCGAATTGAGATGACGCCACTGGCTGGGCGTCATCCCGGTTTC**NNN**GGTAAACACCACCGA	Construct *rhaR* R295X (downstream)
3095	GTGAAGCTTTTAATCTTTCTGCGAATTGAGATGACGCCACTGGCTGGGCGTCATCCCGGT**NNN**CCGGGTAAACACCACCGA	Construct *rhaR* E2965X (downstream)

Oligo Number	Sequence	Use
3096	GTGAAGCTTTTAATCTTTCTGCGAATTGAGATGACGCCACTGGCTGGGCGTCATCCC**NNN**TTCCCGGGTAAACACCACCGA	Construct *rhaR* T297X (downstream)
3097	GTGAAGCTTTTAATCTTTCTGCGAATTGAGATGACGCCACTGGCTGGGCGTCAT**NNN**GGTTTCCCGGGTAAACACCACCGA	Construct *rhaR* G298X (downstream)
3098	GTGAAGCTTTTAATCTTTCTGCGAATTGAGATGACGCCACTGGCTGGGCGT**NNN**CCCGGTTTCCCGGGTAAACACCACCGA	Construct *rhaR* M299X (downstream)
3099	GATGAATTCGGTCACCGCGTGATATTCGCCAGGGACGGG	Construct bHTH2 variants (upstream)
3117	GCGGGATCCTTAAGAAAAACTAAGAGAAGAAGCTATCGATATCG	Clone wild type *virF*-NTD into pDZ1 (downstream)
3118	GCGGGATCCAGAAAAACTAAGAGAAGAAGCTATCGATATCGA	Clone wild type *virF*-NTD into pDZ3 (downstream)
3159	[AminoC6+DY682]TTGCATTGTCTTGTTTTTTTATCTCATTAATC	IR labeled oligo encoding Rns binding site at *cfaA* (top strand)
3160	[AminoC6+DY682]GATTAATGAGATAAAAAAACAAGACAATGCAA	IR labeled oligo encoding Rns binding site at *cfaA* (bottom strand)
3174	GTCCTCTTCAACGTTGCTG**NNN**AAGCTGATTACC	Construct *rhaR* D209X (middle downstream)
3175	GTCCTCTTCACGTTTCGCTGGATGTTGGCGG	Construct *rhaR* D209X (middle upstream)
3176	GTCCTCTTCGGATAAGCTGATT**NNN**CGGCTGGCG	Construct *rhaR* T213X (middle downstream)
3177	GTCCTCTTCTATCCAGCAACGTTTCGCTGGATGTTGG	Construct *rhaR* T213X (middle upstream)
3178	GTCCTCTTCCCGGCTGGCG**NNN**AGCCTGAAAAG	Construct *rhaR* A217X (middle downstream)
3180	GTGCTCTTCATATCTTCTC**NNNTAC**AGCCGCCTGTTAATC	Construct *rhaR* Q268X-H269Y
3181	GTCCTCTTCCAGAGTGTGT**NNN**GCGCAATATCTTC	Construct *rhaR* H262X-H269Y
3182	GTGCTCTTCTCATGCGCAA**NNN**CTTCTCCAG**TAC**AGCCG	Construct *rhaR* Y265X-H269Y
3183	GTGCTCTTCATATCTTCTC**TGCTAC**AGCCGCCTGTTAATC	Construct *rhaR* Q268C-H269Y
3184	GTGCTCTTCATATCTTCTC**CGTTAC**AGCCGCCTGTTAATC	Construct *rhaR* Q268R-H269Y
3185	GTCCTCTTCCAGAGTGTGT**TCT**GCGCAATATCTTC	Construct *rhaR* H262S-H269Y
3186	GTGCTCTTCTCATGCGCAA**AAG**CTTCTCCAG	Construct *rhaR* Y265K-H269Y
3231	GTGCTCTTCGCTGACGCAAAACGCGCTCAC	Construct *rhaR* bHTH1 variants (middle upstream)
3232	GTCCTCTTCTCAGCAA**NNN**CGCCAGCAGACTGGAATGACC	Construct *rhaR* F243X (middle downstream)
3233	GTCCTCTTCTCAGCAATTTCGC**NNN**CAGACTGGAATGACC	Construct *rhaR* Q245X (middle downstream)
3234	GTCCTCTTCTCAGCAATTTCGCCAG**NNN**ACTGGAATGACC	Construct *rhaR* Q246X (middle downstream)
3235	GTCCTCTTCTCAGCAATTTCGCCAGCAG**NNN**GGAATGACC	Construct *rhaR* T247X (middle downstream)
3236	GTCCTCTTCGATATTGATTGATGGTCATTCCAGTCTGCTGG	Construct *rhaR* Q257X or C261X (middle upstream)
3237	GTCCTCTTCATATCTGCGA**NNN**GTCAGAGTGTGTCATGCG	Construct *rhaR* Q257X (middle downstream)

Oligo Number	Sequence	Use
3238	GTCCTCTTCATATCTGCGACAGGTCAGAGTG**NNN**CATGCGC	Construct *rhaR* C261X (middle downstream)
3240	GGTCTCTTCGGAGAAGATATTGCGCAT	Construct *rhaR H*269X (middle upstream) Re-order of 2909
3249	GCGCATATGATGGACATGGGTCACAAAAACAAAATCGACATTAAA	Clone codon optimized *virF* into pDZ1 and pDZ3 (upstream)
3250	GCGGGATCCTTAGAATTTTTTGTGGTACAGGTAGAATTTTTTCGGGG	Clone codon optimized *virF* into pDZ1 (downstream)
3251	GCGGGATCCGAATTTTTTGTGGTACAGGTAGAATTTTTTCGGGGG	Clone codon optimized *virF* into pDZ3 (downstream)
3255	GTGAAGCTTTTAGAATTTTTTGTGGTACAGGTAGAATTTTTTCGGGG	Clone codon optomized *virF* into pHG165 (downstream)
3256	GTGGAATTCTATTCTGTTGAATTTATGGACCAGATTCGTAAAATCGTGG	Clone codon optimized virF-DBD into pHG165 (upstream)
3257	GCGGGATCCATGATGGACATGGGTCACAAAAACAAAATCGACATTAAAG	Clone codon optimized virF into pSE264 (upstream)
3273	GCACATATGAAACTGAAACAGAACATCG	Clone codon optimized aggR into pDZ1 and pDZ3 (upstream)
3274	ACGGGATCCTTATTGCGATTTGAAGTAGGTCAGG	Clone codon optimized aggR into pDZ1 (downstream)
3275	ACGGGATCCTTGCGATTTGAAGTAGGTCAGG	Clone codon optimized aggR into pDZ3 (downstream)
3276	GATGAATTCGATAAAGACATTTTTTCAAGGAGGAAAGATATG	Clone codon optimized aggR into pHG165 (upstream)
3277	GTGAAGCTTTTATTGCGATTTGAAGTAGGTCAGG	Clone codon optimized aggR into pHG165 (downstream)
3278	GCGGGATCCATGAAACTGAAACAGAACATCG	Clone codon optimized aggR into pSE264 (upstream)
3279	GTCCTCTTCGCCGGGTAATCAGCTTATCCAGCAACG	Construct *rhaR* A217X (middle upstream)
3282	GTGGAATTCCGCGTGATATTCGCCAGGGACGGGATGTTAAAA**CGC**CTCAAAGATGATTTTTTTGCC	Construct rhaR Δ2-34 L37R-H269S
3287	GTTCTTACCTTAAATTTTCGACGGAAAACCACGTAAAAAACGTCGATTTTTCAAGATACAGCG	Unlabeled oligo encoding full RhaR binding site at *rhaSR* (btm strand)
3309	GCGGAATTCATGATGGACATGGGTCACAAAAACAAAATCGACATTAAAG	Clone codon optimized *virF* into pMAL-C2 (upstream)
3310	GCGGGATCCATGATGGATATGGGACATAAAAACAAAATAGATATAAAGG	Clone wild type *virF* into pET24b (upstream)
3311	GCGCTCGAGAAATTTTTTATGATATAAGTAAAATTTCTTTGGAG	Clone wild type *virF* into pET24b (downstream)
3313	GCGGAATTCCGCGTGATATTCGCCAGGGACGGGATG**NNN**AAACTTCTCAAAGATGATTTTTTTGCC	Construct rhaR Δ2-34 L35X
3326	TTGCATTGTCTTGTTTTTTTATCTCATTAATC	Unlabeled oligo encoding Rns binding site at *cfaA* (top strand)
3340	GTGGAATTCCGCGTGATATTCGCCAGGGACGGGATGACGGTGGCGCATCAGTTAAAACTTCTC	Construct rhaR Δ2-29
3371	[AminoC6+DY682]GTGCCCTGGTCTGG	Labeling oligo used for LUEGO protocol
3372	AGAATATTATTCTTTTATCCAATAAAGATAAATTGCATCAATCCAGCTATTAAAATAGTA	Oligo encoding full VirF binding site at *virB* (top strand)

Oligo Number	Sequence	Use
3373	TACTATTTTAATAGCTGGATTGATGCAATTTATCTTTATTGGATAAAAGAATAATATTCTCCAGACCAGGGCAC	Oligo encoding full VirF binding site at *virB* (bottom strand)
3374	AGAATATTATTCTTTTATCCAATAAAGA	Oligo encoding half VirF binding site at *virB* (top strand)
3375	TCTTTATTGGATAAAAGAATAATATTCTCCAGACCAGGGCAC	Oligo encoding half VirF binding site at *virB* (bottom strand)
3380	CGATCACTCTTCTCTGTTCGGGCAGTTGG	Middle rhaR (downstream) Splice at L181 (replacement for 2803)
3381	CGATCACTCTTCTCAGCAACTCAGCCATTTC	Middle rhaR (upstream) Splice at L181 (replacement for 2804)
3384	GTCCTCTTCGGATAAGCTGATTACCCGGCTGGCG	Middle rhaR (downstream) Splice at D209
3412	GTGCTCTTCGTTGCTGAATCGCCATCGTTACACCAGTGATTCG	Downstream middle oligo to make RhaR L185E and L181E/L185E.
3413	GTGCTCTTCGCAACATCAC**CTC**CTGCCCGAA**CTC**CAACTCAGC	Upstream middle oligo to make RhaR L181E/L185E
3414	GTGCTCTTCGCAACATCAC**CTC**CTGCCCGAACAGCAACTCAGC	Upstream middle oligo to make RhaR L185E

Underlined, bolded, and italicized nucleotides are not complementary to wild-type DNA sequence. Underlined regions encode restriction endonuclease sites and bolded regions encode mutations. Italicized regions indicate non-native sequence added to the gene.

For oligonucleotides with **NNN, N=** A,G,C,T

cleaned up using the E.Z.N.A. Cycle Pure kit (Omega Bio-Tek, Norcross, GA) or IBI Gel/PCR DNA Fragment Extraction kit (IBI Scientific, Peosta, IA). Transformation was carried out using chemically induced competent cells of *E. coli* and plasmid DNA was purified using either the Plasmid Mini Kit I (Omega Bio-Tek, Norcross, GA) or the High-Speed Plasmid Mini kit (IBI Scientific, Peosta, IA). DNA sequencing was performed at the Northwestern University Genomics Core (Chicago, IL). The DNA sequence of both strands was determined for the entire cloned region of all cloned and mutagenized fragments.

Strains and Plasmids. Tables 2 and 3 contain the list of all strains and plasmids used in this study. All strains used in experiments with RhaR in this study are derivatives of *E. coli* strain ECL116 (151) and all VirF experiments done in *E. coli* in this study are derivatives of *E. coli* strain MC4100 (152), genotypes list additional alleles in these genetic backgrounds. The *S. flexneri* strain used in this study was BS536 (gift from A. T. Maurelli, Uniformed Services University).

All β-galatosidase assays of activation by RhaR or its variants were performed in strain SME3160 [λΦ(*rhaS-lacZ*)Δ85 Δ(*rhaSR*)::Km *recA*::*cat*] (144) (strain SME2999 with *recA*::*cat* introduced by Phage P1-mediated generalized transduction (95)) or SME2525 [λΦ(*rhaS-lacZ*) Δ128 Δ(*rhaSR*)::Km *recA*::*cat*] (31). All β -galatosidase assays of activation by VirF or its variants were performed in strain SME3792 [λΦ(*virB-lacZ*) Δ259]. All *lacZ* fusions used in this study were translational fusions. The construction of new strains for use in this study is described below in the appropriate sections.

Mutagenesis of RhaR Allosteric Site in Subdomain 2 (AS2). Wild-type *rhaR* and all mutants were cloned into and expressed from the plasmid pHG165 (150) using the

Table 2. Strains used in this study

E. coli strains	Genotype	Source or Reference
SME1049	ECL116 [F- Δ*lacU169 endA hsdR thi*]	(151)
SME1541	CSH141	(153)
SME2410	BS536 *Shigella flexneri virB-lacZ*	T. Maurelli
SME2503	ECL116 λϕ(*rhaS-lacZ*) Δ128 *recA*::*cat*	Laboratory collection
SME2525	ECL116 λΦ(*rhaS-lacZ*)Δ128 Δ*rhaSR-kan recA*::*cat*	Laboratory collection
SME3053	ECL116 λϕ(*rhaS-lacZ*) Δ85 *recA*::*cat*	Laboratory collection
SME3160	ECL116 *rhaS-lacZ* Δ85Δ*rhaSR*::*kan recA*::*cat*	Laboratory collection
SME3564	SME3160+ pHG165/*rhaR*	This study
SME3565	SME2525+pHG165/*rhaR*	This study
SME3613	AP76 [[MC4100 λΦ(*nlpA-lacZ*) Δ391]	G.P. Munson
SME3629	SME3792+pATM323/*virF*	T. Maurelli
SME3653	SME3160+pHG165/*rhaR* Y265H	This study
SME3654	SME3160+pHG165/rhaR Y265T	This study
SME3655	SME3160+pHG165/*rhaR* Y265G	This study
SME3656	SME3160+pHG165/*rhaR* Y265L	This study
SME3734	SME3160+pHG165/*rhaR* S270R	This study
SME3735	SME3160+pHG165/*rhaR* S270T	This study
SME3736	SME3160+pHG165/*rhaR* S270D	This study
SME3737	SME3160+pHG165/*rhaR* S270V	This study
SME 3738	SME3160+pHG165/*rhaR* S270I	This study
SME3739	SME3160+pHG165/*rhaR* S270G	This study
SME3750	BL21(DE3) DnaY	R. DeGuzman
SME3754	AF51 [MC4100 λΦ(cfaA-lacZ) Δ469]	G.P. Munson
SME3773	SME3160+pHG165/*rhaR* H262D	This study
SME3774	SME3160+pHG165/*rhaR* H262R	This study
SME3775	SME3160+pHG165/*rhaR* H262V	This study
SME3776	SME3160+pHG165/*rhaR* wt (for N-terminal deletions)	This study
SME3777	SME3160+pHG165/*rhaR* Δ2-34	This study
SME3778	SME3160+pHG165/*rhaR* Δ2-40	This study
SME3792	AB097 [MC4100 λΦ(*virB-lacZ*) Δ259]	G.P. Munson
SME3794	SME3792+pHG165/*virF*-DBD (Δ2-160)	This study
SME3797	SME3160+pHG165/*rhaR* Y265M	This study
SME3798	SME3160+pHG165/*rhaR* Y265E	This study
SME3799	SME3160+pHG165/*rhaR* Y265K	This study

SME3805	Rosetta 2(DE3)	Novagen
SME3821	SME3160+pHG165/*rhaR*-DBD	This study
SME3822	SME3160+pHG165 *Sap*I-/*rhaR*	This study
SME3823	SME3160+pHG165 *Sap*I-/*rhaR*-DBD	This study
SME3824	SME3792+pDZ1/*virF*	This study
SME3825	SME3160+pHG165/*rhaR* Q268C	This study
SME3826	SME3160+pHG165/*rhaR* Q268V	This study
SME3827	SME3160+pHG165/*rhaR* Q268G	This study
SME3828	SME3160+pHG165/*rhaR* Q268M	This study
SME3829	SME2525+pHG165/*rhaR* Δ2-34	This study
SME3830	SME2525+pHG165/*rhaR* Δ2-40	This study
SME3841	SME3792+pDZ3/*virF*-DBD	This study
SME3842	SME3792+pDZ3/*virF*	This study
SME3843	SME3792+pDZ1/*virF*-DBD	This study
SME3850	BL21(DE3)pLysS	Promega
SME3857	SME3160+pHG165/*rhaR* L266V	This study
SME3858	SME2525+pHG165/*rhaR* WT (for N-term deletions)	This study
SME3866	KS1000	New England Biolabs
SME3873	SME3160+pHG165/*rhaR* R295L	This study
SME3874	SME3160+pHG165/*rhaR* R295Q	This study
SME3875	SME3160+pHG165/*rhaR* E296T	This study
SME3876	SME3160+pHG165/*rhaR* E296L	This study
SME3877	SME3160+pHG165/*rhaR* E296Q	This study
SME3878	SME3160+pHG165/*rhaR* E296Y	This study
SME3879	SME3160+pHG165/*rhaR* E296H	This study
SME3880	SME3160+pHG165/*rhaR* T297N	This study
SME3881	SME3160+pHG165/*rhaR* T297A	This study
SME3882	SME3160+pHG165/*rhaR* T297I	This study
SME3883	SME3160+pHG165/*rhaR* T297S	This study
SME3884	SME3160+pHG165/*rhaR* G298L	This study
SME3885	SME3160+pHG165/*rhaR* G298T	This study
SME3886	SME3160+pHG165/*rhaR* G298S	This study
SME3887	SME3160+pHG165/*rhaR* G298K	This study
SME3888	SME3160+pHG165/*rhaR* M299T	This study
SME3889	SME3792+pHG165/*virF*	This study
SME3889	SME3160+pHG165/*rhaR* M299E	This study
SME3890	SME3160+pHG165/*rhaR* M299L	This study

SME3897	SME3862 (SME2999[2985 (S-Z Δ85) ΔrhaSR::kan] malP::lacIq recA::cat) + pHG165rhaR	This study
SME3909	SME3160+pHG165/*rhaR* Q264V	This study
SME3910	SME3160+pHG165/*rhaR* Q264L	This study
SME3910	SME3160+pHG165/*rhaR* R295K	This study
SME3911	SME3160+pHG165/*rhaR* L267F	This study
SME3912	SME3160+pHG165/*rhaR* L267S	This study
SME3913	SME3160+pHG165/*rhaR* L267A	This study
SME3914	SME3160+pHG165/*rhaR* L267M	This study
SME3920	SME3160+pHG165/*rhaR* L267I	This study
SME3921	SME3160+pHG165/*rhaR* L267T	This study
SME3922	SME3160+pHG165/*rhaR* R295I	This study
SME3923	SME3160+pHG165/*rhaR* M299V	This study
SME3924	SME3160+pHG165/*rhaR* H262I	This study
SME3925	SME3160+pHG165/*rhaR* H262S	This study
SME3926	SME3160+pHG165/*rhaR* H262F	This study
SME3927	SME3160+pHG165/*rhaR* H269P	This study
SME3928	SME3160+pHG165/*rhaR* Y265F	This study
SME3930	SME3160+pHG165/*rhaR* H262Y	This study
SME3931	SME3160+pHG165/*rhaR* H262L	This study
SME3932	SME3160+pHG165/*rhaR* Y265P	This study
SME3933	SME3160+pHG165/*rhaR* R271V	This study
SME3934	SME3160+pHG165/*rhaR* L266C	This study
SME3935	SME3160+pHG165/*rhaR* L266E	This study
SME3936	SME3160+pHG165/*rhaR* L266A	This study
SME3937	SME3160+pHG165/*rhaR* H269I	This study
SME3938	SME3160+pHG165/*rhaR* H269L	This study
SME3939	SME3160+pHG165/*rhaR* H269K	This study
SME3940	SME3160+pHG165/*rhaR* H269Y	This study
SME3941	SME3160+pHG165/*rhaR* H269S	This study
SME4117	Arctic Express +pSE290/*rhaR*	Stratagene
SME4158	SME3754+pHG165/aggR^{co}	This study
SME4243	SME2525+ pHG165/*rhaR* Δ34 Q264L	This study
SME4244	SME2525+ pHG165/*rhaR* Δ34 Y265M	This study
SME4245	SME2525+ pHG165/*rhaR* Δ34 L266A	This study
SME4246	SME2525+ pHG165/*rhaR* Δ34 L266E	This study
SME4247	SME2525+ pHG165/*rhaR* Δ34 Q268V	This study
SME4248	SME2525+ pHG165/*rhaR* Δ34 H269Y	This study

SME4249	SME2525+ pHG165/*rhaR* Δ34 H269S	This study
SME4250	SME2525+ pHG165/*rhaR* Δ34 S270G	This study
SME4251	SME2525+ pHG165/*rhaR* Δ34 E296T	This study
SME4252	SME2525+ pHG165/*rhaR* Δ34 T297S	This study
SME4253	SME2525+ pHG165/*rhaR* Δ34 G298S	This study
SME4254	SME2525+pHG165/*rhaR* L37R-H269S	This study
SME4255	SME3160+pHG165/*rhaR* D209A	This study
SME4258	SME3792+pHG165/aggR^{co}	This study
SME4272	SME2525+ pHG165/*rhaR* Δ34 H262V	This study
SME4273	SME2525+ pHG165/*rhaR* Δ34 L267M	This study
SME4274	SME2525+ pHG165/*rhaR* Δ34 R271V	This study
SME4275	SME2525+ pHG165/*rhaR* Δ34 R295Q	This study
SME4276	SME3160+pHG165/*rhaR* Q264K	This study
SME4277	SME3160+pHG165/*rhaR* Q264T	This study
SME4278	SME2525+ pHG165/*rhaR* Δ34 H269T	This study
SME4279	SME2525+ pHG165/*rhaR* Δ34 H269V	This study
SME4280	SME3160+pHG165/*rhaR* Q245C	This study
SME4281	SME3160+pHG165/*rhaR* Q245A	This study
SME4282	SME3160+pHG165/*rhaR* Q245V	This study
SME4283	SME3160+pHG165/*rhaR* Q245G	This study
SME4284	SME3160+pHG165/*rhaR* Q246L	This study
SME4285	SME3160+pHG165/*rhaR* Q246S	This study
SME4286	SME3160+pHG165/*rhaR* F243L	This study
SME4287	SME3160+pHG165/*rhaR* F243M	This study
SME4288	SME3160+pHG165/*rhaR* F243I	This study
SME4289	SME3160+pHG165/*rhaR* Q257I	This study
SME4290	SME3160+pHG165/*rhaR* Q257R	This study
SME4291	SME3160+pHG165/*rhaR* Q257L	This study
SME4292	SME3160+pHG165/*rhaR* Q257V	This study
SME4292	SME3160+pHG165/*rhaR* L35K-H269S	This study
SME4293	SME3160+pHG165/*rhaR* C261T	This study
SME4293	SME3160+pHG165/*rhaR* L35K-H269Y	This study
SME4294	SME3160+pHG165/*rhaR* C261V	This study
SME4294	SME3160+pHG165/*rhaR* L35K-H269V	This study
SME4295	SME3160+pHG165/*rhaR* C261I	This study
SME4296	SME3160+pHG165/*rhaR* Y265Q-H269Y	This study
SME4297	SME3160+pHG165/*rhaR* Y265A-H269Y	This study
SME4298	SME3160+pHG165/*rhaR* Y265G-H269Y	This study

SME4299	SME3160+pHG165/*rhaR* Y265R-H269Y	This study
SME4300	SME3160+pHG165/*rhaR* A217G	This study
SME4301	SME3160+pHG165/*rhaR* A217P	This study
SME4302	SME3160+pHG165/*rhaR* A217V	This study
SME4303	SME3160+pHG165/*rhaR* T247I	This study
SME4304	SME3160+pHG165/*rhaR* T247G	This study
SME4304	SME2525+ pHG165/*rhaR* Δ34 Q268M	This study
SME4305	SME3160+pHG165/*rhaR* T247W	This study
SME4306	SME3160+pHG165/*rhaR* D209E	This study
SME4307	SME3160+pHG165/*rhaR* T213A	This study
SME4308	SME2525+ pHG165/*rhaR* Δ34 H269C	This study
SME4309	SME2525+pHG165/*rhaR* Δ34 Q268M	This study
SME4310	SME3160+pHG165/*rhaR* Q246R	This study
SME4311	SME3160+pHG165/*rhaR* Q246W	This study
SME4312	SME3160+pHG165/*rhaR* T213L	This study
SME4313	SME3160+pHG165/*rhaR* H262Y-H269Y	This study
SME4314	SME3160+pHG165/*rhaR* H262A-H269Y	This study
SME4315	SME3160+pHG165/*rhaR* T247A	This study
SME4319	KS1000+pMAL-C2x/*virF*^{co}	This study
SME4320	SME3160+pHG165/*rhaR* C261L	This study
SME4322	SME2525+pHG165/*rhaR Δ34* L35C-H269S	This study
SME4323	SME2525+pHG165/*rhaR Δ34* L35H-H269S	This study
SME4381	SME3160+pHG165/*rhaR* L37R-H269S	This study
SME4391	SME3160+pHG165/*rhaR* Δ2-29	This study
SME4413	SME2525+pHG165/*rhaR* Δ2-29	This study
SME4419	SME3160+pHG165/*rhaR* W75R-H269K	This study
SME4420	SME3160+pHG165/*rhaR* W75R-H269S	This study
SME4421	SME3160+pHG165/*rhaR* D197G-H269K	This study
SME4422	SME3160+pHG165/*rhaR* L35K-H269K	This study
SME4423	SME3160+pHG165/*rhaR* L37R-H269K	This study
SME4428	SME2503 Tn-10 linked to *pcnB* mutant (from P1 on SME2595)	This study
SME4429	SME2525 Tn-10 linked to *pcnB* mutant (from P1 on SME2595)	This study
SME4430	SME3053 Tn-10 linked to *pcnB* mutant (from P1 on SME2595)	This study
SME4431	SME3160 Tn-10 linked to *pcnB* mutant (from P1 on SME2595)	This study
SME4469	SME3160+pHG165/*rhaR* D41A-T279A	This study
SME4470	SME3160+pHG165/*rhaR* D41S-T279S	This study

Table 3. Plasmids used in this study

Plasmid	Genotype	Source or Reference
pHG165	*lacZα rop* Ap^r (ColE1 origin from pBR322)	(150)
pDZ1	pET21a/GB1 domain-r*unX1T1*	R. DeGuzman
pSE262	pHG165 P*rhaSR*con-2676 (expression vector)	Laboratory collection
pSE263	pHG165 P*rhaSR*con-2689 (expression vector)	Laboratory collection
pSE264	pHG165 P*rhaSR*con-2690 (expression vector)	Laboratory collection
pDZ3	pET21a/runX1T1-GB1 domain	R. DeGuzman
pGEM	*lacZα*	Promega
pHG165 *Sap*I-	pHG165 with *Sap*I site removed	Laboratory collection
pSE290	pDZ1 with $GB1^{basic}$	Laboratory collection
pATM323	pBAD18/*virF*	T. Maurelli
pMAL-C2x	Ap^r encodes N-terminal MBP solubility tag	New England Biolabs

restriction endonucleases *Eco*RI and *Hind*III, and expressed under the control of the *lac* promoter. These plasmids were expressed in derivatives of strain ECL116, which has the *lacI* gene deleted, thus IPTG was not needed to induce expression. Wild type *rhaR* was PCR amplified using oligos 2452 and 2082. Site-specific random mutagenesis of *rhaR* was performed using PCR to make oligonucleotide-directed random mutations (NNN) at each of the desired positions. For each construct, the gene was amplified in two fragments, in most cases with a mutagenized codon in one fragment and no mutations in the other fragment, using the Type IIS restriction endonuclease *Ear*I enabling joining of the fragments without adding a restriction endonuclease recognition sequence to the final *rhaR* sequence (37, 144, 154) upon cloning into pHG165. Wild type *rhaR* was used as template for all PCR amplifications. The specific oligo sets used to create each set of mutants are described in Table 4. The mutagenizing oligonucleotide is indicated with an *:

Table 4. Oligonucleotides to Amplify Fragments to Generate AS2 Variants

Variant	Oligos to amplify upstream DNA fragment	Oligos to amplify downstream DNA fragment
C261X	2452-3236	3238*-2082
H262X	2452-2905	2914*-2082
Q264X	2452-2905	3092*-2082
Y265X	2452-2907	2915*-2082
L266X	2452-2907	2918*-2082
L267X	2452-2907	3093*-2082
Q268X	2452-3037	3014*-2082
H269X	2452-2909	2916*-2082
S270X	2452-2909	2917*-2082
R271X	2452-2954	2955*-2082
L272X	2452-2954	2956*-2082
L273X	2452-2954	2957*-2082

Variants were transformed into strain SME2525 (31), which carries a single-copy fusion of the *rhaSR* promoter (extending upstream to include the RhaR and CRP sites) with the *lacZ*

reporter gene and a deletion of *rhaSR* [λϕ(*rhaS-lacZ*)Δ128 Δ(*rhaSR*)::Km *recA*::*cat*], and plated on Nutrient Agar plates with X-gal, ampicillin and L-rhamnose. Variants with differing levels of activity compared to wild type were picked and streaked to purity, and plasmid DNA was isolated, sequenced, and transformed into strain SME3160. This strain is isogenic with SME2525 except that the fusion of the *rhaSR* promoter with the *lacZ* reporter gene in SME3160 includes the RhaR binding site but not the CRP binding site [λϕ(*rhaS-lacZ*)Δ85 Δ(*rhaSR*)::Km *recA*::*cat*]. The level of transcription activation by wild-type RhaR and the RhaR variants were determined using β-galactosidase assays.

Mutagenesis of RhaR C-terminus of Helix-Turn-Helix 2 (C-HTH2). Mutants were cloned using standard techniques. Random mutations in C-HTH2 were introduced by amplifying *rhaR* using a pair of oligonucleotides with the downstream oligonucleotide carrying a randomized codon (NNN) at the desired position with wild type template. All C-HTH2 constructs were PCR amplified with the upstream oligo 3099. The downstream oligo used to create each set of mutants are as follows: R295X oligo 3094; E296X oligo 3095; T297X oligo 3096; G298X oligo 3097; M299X oligo 3098. Mutants were cloned into pHG165 as described above and transformed into strain SME3160. The level of transcription activation by wild-type RhaR and the RhaR variants were determined using β-galactosidase assays.

Mutagenesis of RhaR Allosteric Site in Subdomain 1 (AS1). Mutants were cloned using the same techniques as used to clone the AS2 variants using wild type template. The specific oligo sets used to create each set of mutants are described in Table 5 and the mutagenizing oligonucleotide is indicated with an *:

Table 5. Oligonucleotides to Amplify Fragments to Generate AS1 Variants

Variant	Oligos to amplify upstream DNA fragment	Oligos to amplify downstream DNA fragment
D209X	2452-3175	3174*-2082
T213X	2452-3177	3176*-2082
A217X	2452-3279	3178*-2082

Activation levels by variants were determined using β -galactosidase assays with variants transformed into strain SME3160.

Mutagenesis of RhaR C-terminus of Helix-Turn-Helix 1 (C-HTH1). Mutants were cloned using the same techniques as used to clone the AS2 variants using wild type template. The specific oligo sets used to create each set of mutants are described in Table 6. The mutagenizing oligonucleotide for each set of oligonucleotides is indicated with an *. Variants were transformed into strain SME3160. The level of transcription activation by wild-type RhaR and the RhaR variants were determined using β-galactosidase assays.

Table 6. Oligonucleotides to Amplify Fragments to Generate C-HTH1 Variants

Variant	Oligos to amplify upstream DNA fragment	Oligos to amplify downstream DNA fragment
F243X	2452-3231	3232*-2082
Q245X	2452-3231	3233*-2082
Q246X	2452-3231	3234*-2082
T247X	2452-3231	3235*-2082
Q257X	2452-3236	3237*-2082

Construction of RhaR Variants to Test for Epistasis. RhaR variants were constructed using PCR amplification to generate fragments of *rhaR* that carried one mutation and were seamlessly pieced together with another fragment carrying the second mutation upon cloning into pHG165. Fragments were cloned from template encoding the corresponding mutation (L35K, L35R, D41A, D41S, W75R or D197G for upstream

fragments; H269S, H269K, T279A or T279S for downstream fragments). All variants were constructed as described in Table 7:

Table 7. Oligonucleotides to Amplify Fragments to Generate RhaR Epistasis Variants

Variant	Oligos to amplify upstream DNA fragment	Oligos to amplify downstream DNA fragment
L35K-H269S	2452-2793	2794-2082
L35K-H269K	2452-3240	2908-2082
L37R-H269S	2452-2793	2794-2082
L37R-H269K	2452-2793	2794-2082
D41A-T279A	2452-2793	2794-2082
D41S-T279S	2452-2793	2794-2082
W75R-H269S	2452-3381	3380-2082
W75R-H269K	2452-3381	3380-2082
D197G-H269K	2452-3381	3380-2082

The control variant T279S was constructed using the following oligo sets: 2452-2793 (template: pHG165/*rhaR* wild type) and 2794-2082 (template: pHG165/*rhaR* T297S). This variant was sub-cloned as the original variant had an N-terminal truncation in addition to the substitution at T279 (Δ24 T279S).

Mutagenesis for RhaR Second-Site Suppressor Screens. For each position of interest (L35, L37, and H269), plasmids encoding all variants with non-wild-type activity (-)rhamnose were pooled as template for PCR amplification and used to generate a library of DNA fragments, with each carrying a single mutation and encoding one domain of RhaR. A second DNA fragment encoding the other domain was generated by PCR using one oligonucleotide that introduced a random substitution (NNN) at the desired position, similar to our approach for generation of single mutants. The two fragments were cloned into pHG165 using the restriction endonuclease *Ear*I to seamlessly join the two DNA fragments (37, 144, 154). This procedure generated a large number of combinations of

mutations, with one mutation within the DNA encoding each domain. Upon ligation, plasmids were transformed into strain SME3160 and variants with activity similar to wild type were picked and streaked to purity. The level of transcription activation by wild-type RhaR and the RhaR variants were determined using β-galactosidase assays. All variants were constructed as described in Table 8. The control variant H269V was constructed using the following oligo sets: 2452-3381 and 3380-2082.

Table 8. Oligonucleotides to Amplify Fragments to Generate RhaR Second-Site Suppressor Variants

Variant	Oligos to amplify upstream DNA fragment	Oligos to amplify downstream DNA fragment
L35K-H269X	2452-3240	2908-2082
L37Var-H269X	2452-3240	2908-2082
L35X-H269Var	2452-2982	2980-2082
L37X-H269Var	2452-2937	2938-2082

Cloning of RhaR N-terminal Deletions. Cloning was performed using standard techniques as previously described. To create the *rhaR* N-terminal deletions, wild-type template was used to amplify the constructs using the downstream oligo 2082 and the following upstream oligos: Δ2-29, 3340; Δ2-34, 3017a; Δ2-40, 3018a. A full-length wild type control with the same upstream sequence as the N-terminal deletions was also constructed using upstream oligo 3019a. Variants were transformed into strains SME2525 and SME3160. The level of transcription activation by wild-type RhaR and the RhaR variants were determined using β-galactosidase assays.

RhaR β-galactosidase Assays. β-galactosidase assays were performed as previously described (155), using the growth protocol of Neidhardt (149) and the assay method of Miller (156). Briefly, all strains for β-galactosidase assays were grown in three

serial steps: tryptone-yeast extract culture with ampicillin; overnight culture (MOPS-buffered minimal medium containing 0.04% glycerol as a limiting carbon source and ampicillin); and growth culture (MOPS-buffered minimal medium with 0.4% glycerol, ampicillin, and with or without 0.4% L-rhamnose). Activities were averaged from two or three independent experiments with two replicates in each experiment. The activities of variants were considered significantly different from wild type if they had an approximate 1.5-fold change in activity compared with wild type and non-overlapping standard deviation confidence intervals. We were interested in variants that differentially affected RhaR activity (-) versus (+)rhamnose. Thus, we separately compared the (-)rhamnose β-galactosidase activities of variants to that of wild type (-)rhamnose, and the (+)rhamnose activities to wild type (+)rhamnose. Percent of wild-type activation was calculated by dividing the activity of each variant, either (-) or (+)rhamnose, by the wild-type value (-) or (+)rhamnose, respectively, and multiplying by 100.

For assays to test the effect of protein levels on (-)rhamnose activity, pHG165*rhaR* was transformed into strain SME3897, which is isogenic to SME3160, but with a *malP::lacI*^{q} allele introduced to allow controlled expression from the *lac* promoter. Cells were grown as described above but were grown to an OD_{600} of 0.1 in growth culture, placed on ice for approximately 30 minutes, and then expression of *rhaR* induced with the specified amount of IPTG. Induction was static at 37°C for three hours. Although this was the protocol used for this experiment, cells should not have been placed on ice and induction should have been performed with aeration.

RhaR Western Blots. Cultures were grown the same as for β-galactosidase assays. After growth to mid-log in growth media with or without L-rhamnose, 1 mL of cells was

pelleted and re-suspended in 45 µL of SDS running buffer (0.191M glycine, 0.0247M Tris base, 0.1% sodium dodecyl sulfate) with 5 µL loading dye (0.6M Tris base, 2.0% sodium dodecyl sulfate, 0.7M 2-mercaptoethanol, 5.0% glycerol, 0.2% bromophenol blue). For each sample, 20µl was then loaded onto a 12% SDS-polyacrylamide gel, electrophoresed, and blotted onto a nitrocellulose membrane using standard procedures. Primary antibodies against RhaR were custom-made polyclonal rabbit antibodies from Cocalico Biologicals (Reamstown, PA) (93). Anti-RhaR antibodies were affinity purified against N-terminal His_6-tagged full-length RhaR as specified below. Alexa Fluor 680-labeled secondary antibody (anti-rabbit) was obtained from Molecular Probes (Eugene, OR). Primary antibodies against DnaK were commercially available mouse monoclonal antibodies used to normalize protein levels (Abcam, Cambridge, MA). IRD800-labled secondary antibodies (anti-mouse) were purchased from LI-CORE (Lincoln, NE). Blots were imaged using an Odyssey Infrared Imaging System (LI-COR, Lincoln, NE) and quantified using Image J 1.46 software (free download rsbweb.nih.gov).

Affinity-Purification of Anti-RhaR Antibody. RhaR-His_6 protein was expressed from pET24b and purified as previously described for the His_6-RhaR DBD protein (93) but using AKTAexplorer FPLC (GE Healthcare) using a 5 mL HiTrap Chelating HP column (GE Healthcare). The HiTrap Chelating HP column had been charged with 50 mM $NiSO_4$, and equilibrated with 15 mL H_20 and then 15 mL binding buffer (5mM imidazole, 0.5M sodium chloride, 20mM Tris-HCl [pH 7.9]) with 6M urea. After loading, a step gradient of binding buffer containing 6M urea with 20mM to 500 mM imidazole from was run with 20ml volume passed over the column at each concentration. Each step had an increase of 10mM imidazole for the first 100mL with 75mM, 200mM, and 500mM imidazole for the last 3

steps. Approximately twenty micrograms of unfolded protein was then loaded onto four 15% SDS-polyacrylamide gels, electrophoresed, and blotted onto a nitrocellulose membrane using standard procedures. The protein bands were then cut out from the membrane, blocked for two hours, and incubated over night at 4°C with 0.5mL raw serum containing anti-RhaR antibodies in 10mL phosphate buffered saline. The membranes were washed twice with PBS/Tween (0.1% Tween-20) and once with PBS. Anti-RhaR antibodies were released from the membrane by rocking vigorously at room temperature with 3ml of 100mM glycine (pH 2.5) for fifteen minutes. The suspension was then neutralized with one-tenth volume of 1M Tris (pH 8.0).

Generalized Transduction with Tn10-*pcnB* to Reduce Copy Number of the pHG165 Plasmid. A generalized P1 transduction was used to move Tn10-*pcnB* into strains SME2503, 2525, 3053, and 3160. Transductants were identified using a delayed selection on nutrient agar containing tetracycline and sodium pyrophosphate (1.25mM). Re-streaking candidates on nutrient agar containing tetracycline confirmed transduction of the Tn10 marker. Transductants of strains SME2525 (SME4429) and SME3160 (SME4431) were then transformed with pHG165/*rhaR* to test for activation at *rhaS-lacZ* compared to activation of *rhaS-lacZ* by chromosomal *rhaR* in strains SME4428 and SME4430 transformed with empty vector.

Purification of GB1^{basic}-RhaR. Protein was expressed in *Escherichia* coli and purified by Ni^{2+}-affinity chromatography (148). Briefly, plasmid pSE290 derived from pET21 and expressing GB1^{b}-RhaR was transformed into competent cells of strain ArcticExpress(DE3) (Agilent Technologies, Santa Clara CA). The cells were grown in 1 liter TY plus ampicillin and gentamycin plus rhamnose. The cells were grown to OD_{600} of 0.5,

transferred to a 15°C shaker, 0.1 mM IPTG was added, and then incubated overnight. Cells were harvested by centrifugation and then resuspended in 30 mL of cold binding buffer (20 mM Tris, 500 mM NaCl, 5 mM imidazole, pH 7.9) plus L-rhamnose. Cells were lysed by three cycles of freeze thaw [with addition of lysozyme (0.4 mg/mL), tris(2-carboxyethyl) phosphine (TCEP, 1 mM) and phenylmethylsulfonyl fluoride (PMSF, 1 mM), at -80°C] followed by sonication, and then centrifuged to remove cell debris. The supernatant was applied using an AKTAexplorer FPLC (GE Healthcare) to a 5 mL HiTrap Chelating HP column (GE Healthcare) that had been charged with 50 mM $NiSO_4$, and equilibrated with 15 mL H_20 and then 15 mL binding buffer. After loading, the column was washed with 25 mL binding buffer, then 25 mL wash buffer (binding buffer, but with 60 mM imidazole) plus L-rhamnose. A 10 mL gradient of binding buffer with 60 mM to 250 mM imidazole from was run and then 15 mL of elution buffer (binding buffer, but with 250 mM imidazole). The ArcticExpress cold-adapted chaperonins Cpn10 and Cpn60 (14 monomers per unit) co-purified with $GB1^b$-RhaR, thus $GB1^b$-RhaR represented only approximately 20% of the total protein used in the assays.

RhaR *In vitro* DNA Binding Assays. Electrophoretic mobility shift assays were performed as described (31), with the following modifications. Reaction volumes were 12 μL total (with 5 μL loaded in each lane), in 1x EMSA buffer [10 mM Tris-HCl (pH 7.4), 1 mM KEDTA, 50 mM KCl, 1 mM dithiothreitol, 5% (v/v) glycerol, 0.1 mg/mL bovine serum albumin (BSA) and 500 ng salmon sperm DNA]. Reactions also contained additives Nonidet P40 and L-rhamnose. Purified proteins were buffer exchanged into 1x EMSA buffer minus BSA and salmon sperm DNA, and without the addition of additives. Electrophoresis was performed in 0.25× TBE (final concentrations: 22.25 mM Tris base,

22.25 mM boric acid, 500 µM disodium EDTA, pH 8.3). All EMSA reactions, including those without inhibitor, had a final concentration of 10% DMSO (SE-1 solvent). DNA probes were generated by hybridizing the following oligonucleotides (oligos): oligo 3056 (5'-[IRD700]CGCTGTATCTTGAAAAATCGACGTTTTTTACGTGGTTTTCCGTCGAAAATTTAAG GTAAGAAC-3') and oligo 3287 (5'-GTTCTTACCTTAAATTTTCGACGGAAAACCACGTAAAAA ACGTCGATTTTTCAAGATACAGCG-3'). IRD700 labeled oligos were from Eurofins MWG Operon. For each oligo pair, 100 µmol of each oligo was combined and the reaction was diluted in STE buffer (50 mM NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) to 20 µL, heated to 94°C for 2 min, and cooled to room temperature. The double-stranded DNA probes were further diluted in STE, and 0.3µL added to EMSA reactions. EMSA gels were imaged using an Odyssey infrared imager (LI-COR, Lincoln, NE), and quantified using the Odyssey software, version 3.0.30. Error bars represent the standard error of the mean. Inhibition values were calculated and graphs were drawn as for *in vivo* dose-response experiments.

Cloning of *virF*. Wild type *virF* was moved from plasmid pATM323 into pHG165 using oligos 2958 and 3042. A Δ2-160 construct was also generated using oligos 2968 and 3042. The wild type gene was also cloned into expression vectors pDZ1 and pDZ3 (gift from Dr. Roberto DeGuzman) to generate N- and C-terminal tagged fusion proteins with GB1 using upstream oligo 3048 with either 3049 (pDZ1) or 3050 (pDZ3) for the downstream oligo. The DNA Binding Domain (DBD) of VirF was also cloned into pDZ1 and pDZ3 using upstream oligo 3068 with either 3049 or 3050, respectively. The wild-type *virF* gene was also cloned into the expression vector pET24b to generate a C-terminal His6-fusion protein using oligos 3310 and 3311.

Cloning of *virF*^{co}. A codon-optimized version of *virF* (*virF*^{co}) was ordered from

GenScript (Piscataway, NJ). The codon optimized gene was cloned into the pMAL-C2x expression vector (New England Biolabs Beverly, MA) to generate an N-terminally tagged fusion protein with Maltose Binding Protein (MBP) using oligos 3309 and 3255. The codon optimized gene was also cloned into pSE290 to generate an N-terminally tagged fusion protein with $GB1^{basic}$ using oligos 3249 and 3250.

VirF β-galactosidase Assays. β-galactosidase assays were performed as previously described (155), using the Miller method (156). Briefly, all strains for β-galactosidase assays were grown in the following three serial steps: tryptone-yeast extract culture with ampicillin; overnight culture (MOPS-buffered minimal medium containing 0.04% glycerol as a limiting carbon source and ampicillin); and growth culture (MOPS-buffered minimal medium with 0.4% glycerol as the carbon source, with ampicillin) based on the method of Neidhardt (149). Strains grown in LB for β-galactosidase assays were grown overnight in LB and then sub-cultured to fresh media. For cultures grown β-galactosidase assays to test potential effectors, growth culture also contained candidate effectors at the concentrations specified in individual assays. For most assays, specific activities were averaged from at least two, usually three, independent assays with two replicates in each assay. Every assay performed included two replicates. β-galactosidase activity is expressed in Miller Units.

Screen for Expression of VirF with Solubility Tags. Protein expressed from either pDZ1 or pDZ3 was transformed into *E. coli* BL21 (DE3) DnaY, BL21 (DE3) pLysS, or BL21 (DE3) Rosetta2 cells as described below. Briefly, plasmids pDZ1 and pDZ3 expressing either full-length *virF* or the VirF DBD were chemically transformed into competent cells of strains BL21 (DE3) DnaY (gift, Roberto DeGuzman), BL21 (DE3) pLysS (Promega Madison, WI), or Rosetta2 (DE3) (Novagen Billerica, MA). Transformants were

grown overnight in TY broth with either kanamycin or chloramphenicol and overnight growth was sub-cultured into 10ml TY with kanamycin and ampicillin, or chloramphenicol and ampicillin and grown to an ~OD_{600} of 0.8. Cultures were transferred to a 15°C shaker, 1mM IPTG was added and then incubated overnight. Cells were harvested by centrifugation and re-suspended in 30ml of cold phosphate binding buffer (20mM potassium phosphate, 0.5M NaCl, 5mM imidazole, pH 7.8). Cells were lysed by three cycles of freeze thaw [with addition of lysozyme (0.4 mg/mL), tris(2-carboxyethyl)phosphine (TCEP, 1 mM) and phenylmethylsulfonyl fluoride (PMSF, 1 mM)] followed by sonication and centrifuged to remove cell debris. Supernatant was electrophoresed on a 12% SDS-polyacrylamide gel and stained with Commassie Blue to assess expression of the GB1 tagged VirF proteins.

Plasmid pET24b/*virF* was transformed into competent cells of strain BL21 (DE3) pLysS (Promega Madison, WI) as previously described (108). Briefly, transformants were grown overnight in TY broth with kanamycin and chloramphenicol. Overnight growth was sub-cultured into 10mL TY with kanamycin and chloramphenicol and grown to OD_{600} of 0.8. Expression was induced with 1mM IPTG at 37°C for three hours. Culture was centrifuged and the pellet re-suspended in 1ml of SDS buffer (0.191M Glycine, 0.0247M Tris base, 0.1% Sodium dodecyl sulfate) and boiled for ten minutes to lyse cells. Samples were electrophoresed on a 12% SDS-polyacrylamide gel and stained with Commassie Blue.

Toxicity Testing of GB1-tagged VirF. Toxicity testing was performed per the pET Systems Manual (Novagen, Billerica, MA). Briefly, VirF protein was expressed from either pDZ1/*virF* or pDZ3/*virF* in *E. coli* SME3750 cells. Transformants were grown in TY with ampicillin and kanamycin overnight. Cultures were sub-cultured to 10ml TY with

ampicillin and kanamycin and grown to ~OD_{600}=0.6. Serial dilutions of 10^{-5} and 10^{-6} were made. For the 10^{-5} dilution, 1ml was then plated onto the following plates: nutrient agar with kanamycin and IPTG (1mM) and nutrient agar with kanamycin, ampicillin, and IPTG (1mM). One milliliter of the 10^{-6} dilution was also plated onto nutrient agar with kanamycin, and nutrient agar with kanamycin and ampicillin. Plates were incubated at 37°C overnight.

Purification of MBP-VirF^{co}. Protein was expressed in *Escherichia coli* and purified by Amylose affinity chromatography as previously described (147). Briefly, plasmid pMAL-C2x expressing MBP-VirF^{co} was transformed into competent cells of strain KS1000 (New England Biolabs Beverly, MA). The cells were grown in 100mL TY plus ampicillin. The cells were grown to OD_{600} of 0.5, transferred to a 15°C shaker, 0.1 mM IPTG was added, and then incubated overnight. Cells were harvested by centrifugation and then re-suspended in 10 mL of cold binding buffer (20 mM Tris, 500 mM NaCl, 1mM EDTA 1mM DTT, pH 7.4). Cells were lysed by three cycles of freeze thaw [with addition of lysozyme (0.4 mg/mL), tris(2-carboxyethyl)phosphine (TCEP, 1 mM) and phenylmethylsulfonyl fluoride (PMSF, 1 mM), at -80˚C] followed by sonication, and then centrifuged to remove cell debris. The supernatant was applied using a BioLogic LP (Bio-Rad Hercules, CA) to 5 mL Amylose Resin (New England Biolabs) that had been equilibrated with 80 mL binding buffer. After loading, the column was washed with 120 mL binding buffer. MBP-VirF^{co} was eluted with elution buffer (binding buffer with 10mM maltose).

Gel Filtration of MBP-VirF^{co}. Gel filtration was performed using a Superdex200 10/300 GL column attached to an AKTAexplorer FPLC (GE Healthcare) equilibrated with 24ml of water and 48ml Elution Buffer. After equilibration, 0.44mL of purified protein was

injected using a 0.44mL loop. Size exclusion chromatography was performed for one column volume (24mL) at a steady flow rate of 0.25mL/min. Prepared gel filtration standards (Bio-Rad Hercules, CA) diluted in Elution Buffer were applied to the column to determine the elution profile of five known proteins. Dimer and aggregate amounts were calculated as a percentage of total areas for all protein peaks in the chromatogram.

VirF *In Vitro* DNA Binding Assays. Electrophoretic mobility shift assays were performed as described (147, 157), with the following modifications. Briefly, reaction volumes were 12 µL total (with 5 µL loaded in each lane), in 1x EMSA buffer [10 mM Tris-HCl (pH 7.4), 1 mM KEDTA, 50 mM KCl, 1 mM dithiothreitol, 5% (v/v) glycerol, 0.1 mg/mL bovine serum albumin (BSA) and 500 ng salmon sperm DNA]. Electrophoresis was performed in 0.25× TBE (final concentrations: 22.25 mM Tris base, 22.25 mM boric acid, 500 µM disodium EDTA, pH 8.3). Experiments with inhibitor, including those reactions without inhibitor, had a final concentration of 10% DMSO (SE-1 solvent). DNA probes were generated by hybridizing the following oligonucleotides (oligos) using the LEUGO procedure (158): For Full length *virB* binding site oligo 3371, 3372, and 3373; for single *virB* binding site oligos 3371, 3374, and 3375. DNA probes encoding the *cfaA* binding site were generated by hybridizing oligonucleotides 3160 with 3159 or 3326. DY682-labeled oligo was purchased from Eurofins MWG Operon. For each oligo pair, 100 µmol of each oligo was combined and the reaction was diluted in STE buffer (50 mM NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) to 20 µL, heated to 94°C for 2 min, and cooled to room temperature. The double-stranded DNA probes were further diluted in STE, and 0.3µL added to EMSA reactions. EMSA gels were imaged using an Odyssey infrared imager (LI-COR, Lincoln, NE), and quantified using the Odyssey software, version 3.0.30. Error

bars represent the standard error of the mean. Inhibition values were calculated and graphs were drawn as for *in vivo* dose-response experiments.

***In Silico* Analysis of RhaR and VirF Protein Structures**. Full length RhaR protein sequence was uploaded to the I-TASSER (Iterative Threading Assembly Refinement) server at http://zhanglab.ccmb.med.umich.edu/I-TASSER for structural predictions using the structure of the apo form of the AraC dimerization domain (PDB ID 1XJA) as a restraint. Full-length VirF protein sequence was also uploaded to I-TASSER using standard methods. I-TASSER utilizes multiple-threading alignments and iterative structural assembly simulations to generate three-dimensional (3D) atomic models from protein sequences (70, 159, 160). Models from I-TASSER were visualized using the MacPyMOL Molecular Graphics System, Version 1.3 (Schrödinger, LLC).

Chapter III: Mechanism of Allosteric Signaling in RhaR

Previous studies of the RhaR protein have led us to propose a model in which L-rhamnose induced changes in interdomain contacts increase both DNA binding and transcription activation (90, 94). Transcription activation presumably occurs through modulation of RhaR contacts with RNAP σ^{70} and/or the C-terminal domain of the RNAP α-subunit (37, 95). RhaR residue T279 is positioned between two known σ^{70} contacts and likely participates in the L-rhamnose-dependent modulation of contacts with RNAP (144) (Figure 6).

In this work, we investigated the role of residues in four regions of the RhaR DBD that were predicted to make interdomain contacts in allosteric signaling: allosteric site in subdomain 1 (AS1); allosteric site in subdomain 2 (AS2); C-terminal HTH1 (C-HTH1); and C-terminal HTH2 (C-HTH2) (Figures 6 and 7). In total, 31 residues within these four regions of the RhaR DBD were subjected to site-directed random mutagenesis (Figures 6 and 7). Pools of mutants were transformed into strain SME2525 to screen for differing levels of activity compared to wild-type RhaR. For each position, ten variants that were representative of the differing activities were chosen in ratios relative to the frequency of each phenotype following transformation. These variants were then streaked to purity, the plasmid DNA isolated and sequenced. Variants with single substitutions were then transformed into strain SME3160 for further study.

In the simplest case, residues involved in RhaR allosteric L-rhamnose signaling would be expected to be involved in one of the following: stimulatory interdomain contacts that increase RhaR activity (+)rhamnose, or inhibitory interdomain contacts that decrease RhaR activity (-)rhamnose. Therefore, one of two outcomes would be expected at any

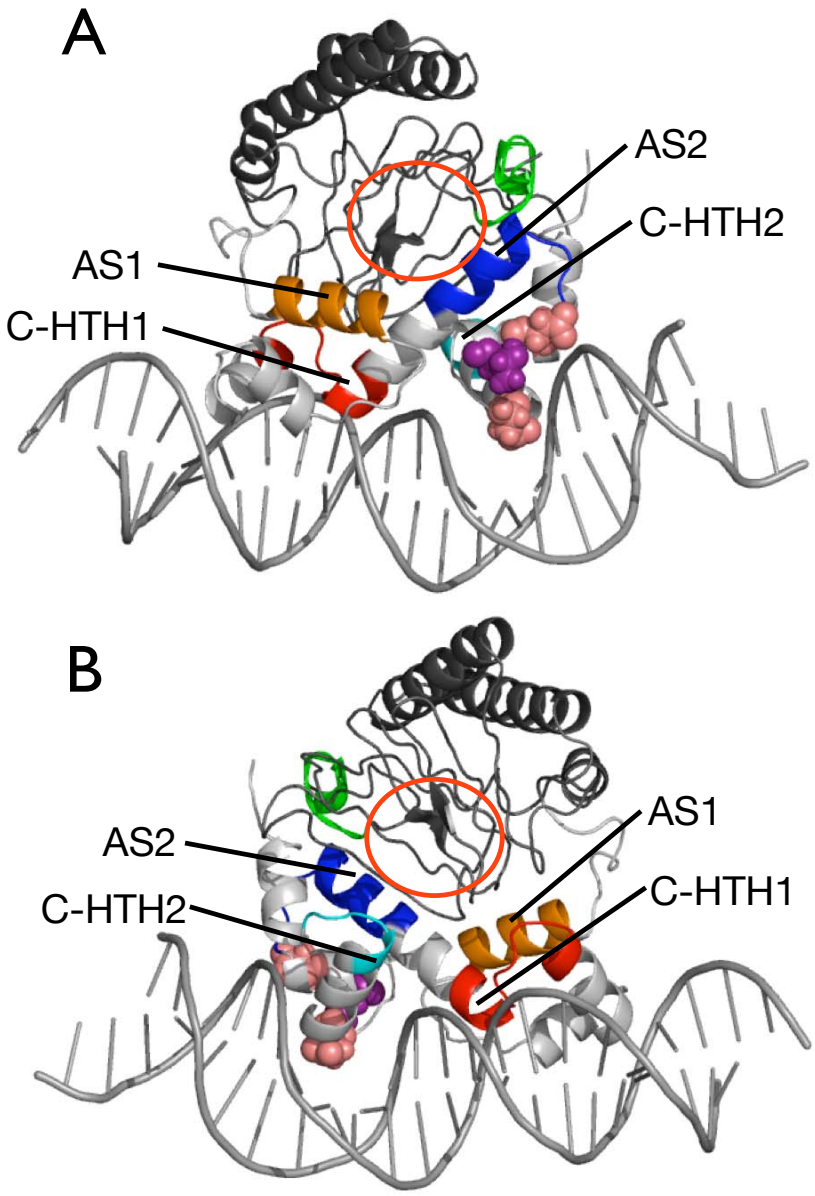


Figure 6. *In Silico* Prediction of the RhaR Structure. Predicted structure of RhaR (-)rhamnose modeled onto DNA from the MarA co-crystal (43). The RhaR NTD is colored dark gray, the DBD is colored light gray. σ^{70} contacting residues are shown as light pink spheres. Residue T279 is shown as a purple sphere. Red circles indicate the predicted L-rhamnose binding pocket. Regions of the RhaR protein that are discussed in this work are highlighted as follows: Arm in green; AS1 in orange; C-HTH1 in red; AS2 in blue; C-HTH2 in cyan. A. "Front" view of RhaR. RNAP would be positioned to the right of the molecule. B. "Back" view of RhaR. RhaR is rotated 180° relative to the "Front" view.

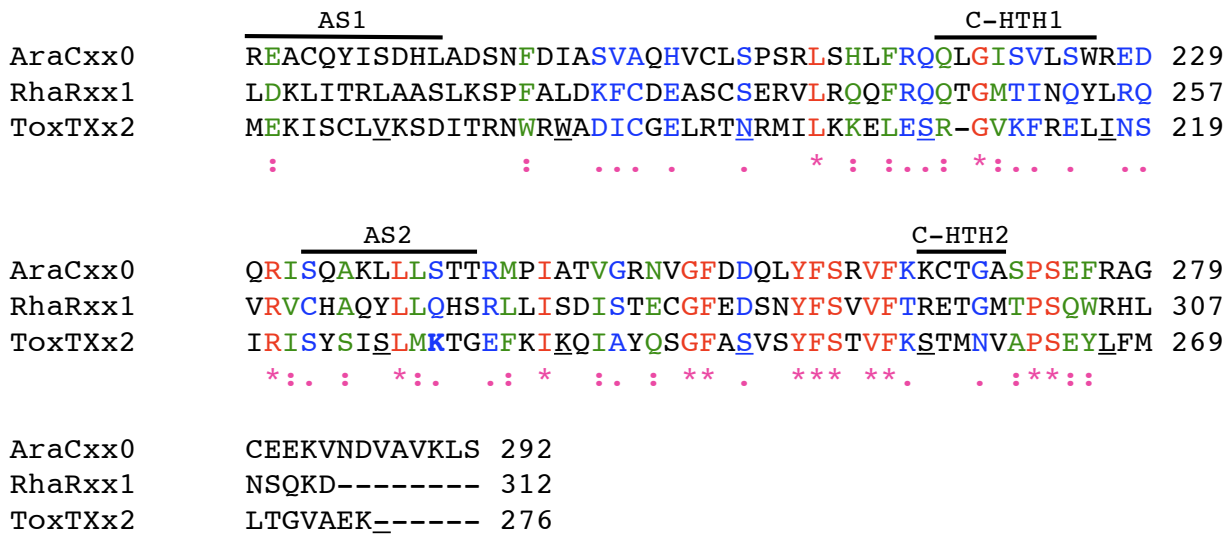


Figure 7. Sequence Alignment of the RhaR, AraC and ToxT DBD's. Amino acid sequences of the RhaR, AraC and ToxT DBD's were aligned using the ClustalW2 program on the EMBL-EBI server. Regions of the RhaR DBD predicted to make interdomain contacts are indicated by bars above the sequences. Every tenth residue (relative to the AraC sequence) is underlined. A (*) below a residue indicates identity, (:) indicates strong similarity, and (.) indicates weak similarity.

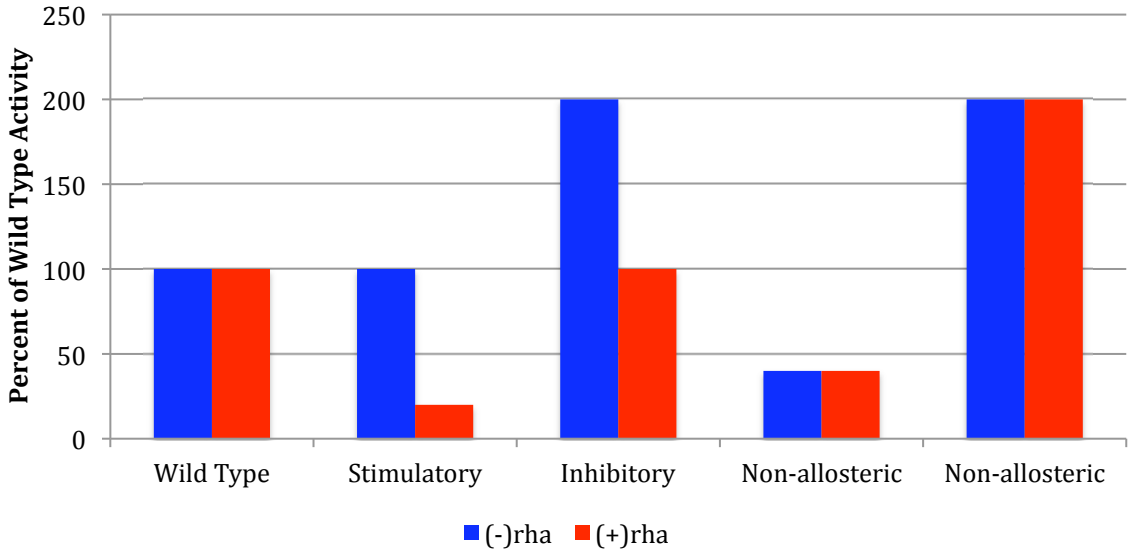


Figure 8. Predicted Outcomes for Variants Involved in Allosteric Contacts. At a given position, loss of function (at the residue level, not whole protein level) variants involved in stimulatory contacts would have decreased activity only (+)rhamnose. Loss of function variants involved in inhibitory contacts would have increased activity only (-)rhamnose. Variants equally affecting both states are not predicted to be involved in L-rhamnose allosteric signaling.

given position: for stimulatory contacts, loss of function (at the residue level, not whole protein level) variants would have decreased activity (+)rhamnose, but wild-type activity (-)rhamnose; for inhibitory contacts, loss of residue function variants would have increased activity (-)rhamnose, but wild-type activity (+)rhamnose. Variants that equally affect both states are unlikely to be involved in L-rhamnose allosteric signaling (Figure 8).

Site-Specific Random Mutagenesis of RhaR AS2. We hypothesized that a group of residues in the RhaR DBD that we refer to as allosteric site in subdomain 2 (AS2) (Figure 6) may make interdomain contacts with the NTD that are involved in allosteric signaling. The RhaR AS2 region aligns with a group of AraC DBD residues (in helix-4 of the DBD, on the DNA-distal surface of the DBD) (Figure 7) that are contacted by the AraC Arm in the absence of the effector arabinose to inhibit transcription of *araBAD* by forming a DNA loop (17, 20, 21, 23-26). RhaR AS2 also aligns with some of the residues in the ToxT DBD that enclose part of the effector-binding pocket in the ToxT NTD and in the presence of effector are positioned to contact the DBD (79, 161), further suggesting that they may be involved in effector-dependent interdomain contacts responsible for transmission of the effector-binding status. We therefore hypothesized that residues in RhaR AS2 may be involved in RhaR allosteric L-rhamnose signaling.

Here, I investigated whether the RhaR AS2 regions plays a role in transmission of the allosteric signal that converts RhaR from its (-)rhamnose (non-activating) state to its (+)rhamnose (activating) state. I randomly mutagenized one codon at a time in the RhaR AS2 region to identify residues that may be involved in this L-rhamnose-dependent signal transmission. 9 positions in the RhaR AS2 region were mutagenized and 51 unique variants were identified. These were assayed for transcription activation in both the absence and

presence of L-rhamnose using a fusion of the *rhaSR* promoter with *lacZ* [(*rhaS-lacZ*)Δ85] (Table 9). Results indicated that residues in the AS2 region of RhaR may impact overall protein stability and the level of active RhaR protein in the cell. Additionally, variants at position 269 support a hypothesis that this residue may be involved in inhibitory contacts that inhibit the activity of RhaR (-)rhamnose.

Of the variants isolated in the RhaR AS2 region, 30% of the substitutions had decreased activity (-) and (+)rhamnose (Table 9, Down). Given that AS2 is predicted to be within one of the seven α-helices that make up the RhaR DBD, it is not surprising that variants in AS2 might have an impact on the overall structure/ function of RhaR. Residues aligning with L267, where all of the variants fall into the Down category, are buried in the structures of MarA, Rob, ToxT, and the AraC DBD, suggesting this residue may have a role in stabilizing the hydrophobic core of the DBD (43, 72, 79, 162). The four variants at Q268 exhibited wild-type activity, indicating that Q268 may not be important for the structure or function of RhaR, despite the fact that the aligned residue in ToxT contacts the NTD effector-binding pocket (79). Alternatively, additional substitutions at position 268 may have been toxic. Notably, and similar to the region of AraC that aligns with AS2 (68), none of the variants in RhaR AS2 had decreased activity (+)rhamnose without also decreasing (-)rhamnose activity, indicating this region is unlikely to function to receive a stimulatory signal from the NTD in the presence of L-rhamnose.

The largest class of variants in RhaR AS2 (50%) had decreased activity (-)rhamnose but approximately wild-type activity (+)rhamnose (Table 9, Down (-)). Protein levels were measured by Western blots for 7 of the 26 variants with this phenotype (C261I, H262L, Q264L, Y265K, L266C, H269I, S270G) (Table 9). The protein levels in cells grown

Table 9. Transcription Activation of *rhaS-lacZ* and RhaR AS2 Variants[a].

Mutation	(-) Rha		(+) Rha		Category[c]
	% WT Activation[a]	% WT Protein Levels[b]	% WT Activation[a]	% WT Protein Levels[b]	
WT	100±25	100	100±13	100	
C261I	22±4	110±39	113±5	93±10	Down (-)
C261V	29±3		77±2		Down (-)
C261L	29±2		86±9		Down (-)
C261T	40±3		95±3		Down (-)
H262D	8±1	57±8	15±1	68±14	Down
H262V	8±1		25±5		Down
H262I	8±1		5±0.2		Down
H262Y	13±3		47±11		Down
H262L	10±0.3	56±4	153±50	51±29	Down (-)
H262S	11±2		117±14		Down (-)
H262F	13±0.2		88±10		Down (-)
H262R	86±7		139±14		~WT
Q264T	21±2		104±8		Down (-)
Q264L	27±4	73±14	93±2	84±5	Down (-)
Q264K	35±3		102±17		Down (-)
Q264V	36±5		111±5		Down (-)
Y265T	9±1		85±3		Down (-)
Y265G	11±1		95±9		Down (-)
Y265K	19±1	51±11	102±21	79±9	Down (-)
Y265E	30±13		119±4		Down (-)
Y265M	32±3		121±20		Down (-)
Y265L	9±0.2		155±6		Down (-) Up (+)
Y265F	81±6		110±17		~WT
Y265H	124±4		120±6		~WT
L266E	9±1		27±5		Down
L266A	8±0.1		82±3		Down (-)
L266C	9±0.4	45±1	86±4	70±11	Down (-)
L266V	11±3		93±10		Down (-)
L267S	8±1		2±0.1		Down
L267F	9±1		9±1		Down
L267A	9±0.5	52±23	11±1	71±9	Down
L267T	10±1		5±0.3		Down
L267M	11±2	53±13	56±2	79±5	Down
L267I	13±4		58±4		Down
Q268V	71±11		77±1		~WT
Q268G	86±10	76±18	70±5	164±97	~WT
Q268C	99±3		95±12		~WT
Q268M	106±21		77±4		~WT

H269P	11±0.4		2±0.1		Down
H269L	12±2		78±2		Down (-)
H269I	17±0.5	57±25	94±20	79±2	Down (-)
H269Y	25±2		102±12		Down (-)
H269S	184±34	78±30	94±5	162±85	Up (-)
H269K	185±8	75±24	111±2	150±76	Up (-)
S270V	10±1		15±1		Down
S270I	10±0.2		3±0.2		Down
S270D	14±1		86±4		Down (-)
S270R	15±1		101±2		Down (-)
S270G	20±2	38±23	106±4	85±28	Down (-)
S270T	116±8		106±6		~WT

a. Variants were assayed in groups with a wild type RhaR activity range of 1.6-2.3 Miller Units (-) Rha and 8.1-10.0 Miller Units (+) Rha. Values are the average of at least two independent assays with two replicates each and are shown as the percent of the corresponding (-) Rha or (+) Rha wild type values. Error is shown as the standard deviation converted to percent of the Miller Unit values.

b. Quantification of variant protein levels compared to wild-type RhaR. Total protein was separated by 12% SDS-PAGE. Proteins were transferred to nitrocellulose, probed with anti-RhaR and anti-DnaK antibodies and quantified using densitometry analysis. RhaR values were normalized to DnaK levels, are the average from two replicates and are shown as the percent of protein relative to the corresponding (-) Rha or (+) Rha wild type values. Error is shown as the standard deviation converted to percent of the normalized protein values.

c. Variants were categorized as follows: Down, reduced activity (-) and (+) Rha; Down (-), reduced activity (-) Rha, near wild-type (+) Rha; ~WT, near wild-type activity (-) and (+) Rha; Up (-), increased activity (-) Rha, wild-type activity (+) Rha; Up, increased activity (-) Rha and (+) Rha.

(-)rhamnose did not correlate with variant activity (-)rhamnose. The (+)rhamnose protein levels were wild type except for H262L, which had two-fold lower protein levels both in the presence and absence of L-rhamnose. Although the amount of protein was not greatly decreased for the variants tested, we cannot exclude the possibility that the substitutions decreased the amount of functional RhaR protein, perhaps by influencing protein folding. Therefore, we hypothesized that decreases in RhaR stability or protein levels may have a greater impact on RhaR (-)rhamnose than (+)rhamnose.

To test the hypothesis that the RhaR protein is more susceptible to changes in protein levels or stability (-)rhamnose than is the protein (+)rhamnose, RhaR was expressed from pHG165 under the control of the *lac* promoter in a strain carrying the *lacI*^{q} allele to the modify transcription levels to different extents with varying concentrations of IPTG (uninduced, 0.1 mM, or 1.0 mM) (Figure 9). We observed a concentration dependent decrease in activity in the (-)rhamnose state while activity (+)rhamnose did not decrease until protein concentrations decreased below 18% of the wild-type level. This could be explained by the protein levels falling in a range where the amount of RhaR binding at *rhaSR* is saturated in the (+)rhamnose state, but not in the (-)rhamnose state. The difference in saturation levels between the (-) and (+)rhamnose states is likely due to the approximate 20-fold difference in DNA binding affinity between the two states (94). Given this difference, in the (-)rhamnose state higher protein levels are required than in the (+)rhamnose state to saturate binding.

Protein levels were similar (-) and (+)rhamnose for both uninduced controls (with and without the *lacI*^{q} allele), which allowed us to determine the maximal effect that can be attributed solely to stability. The difference in activation between the highest and lowest

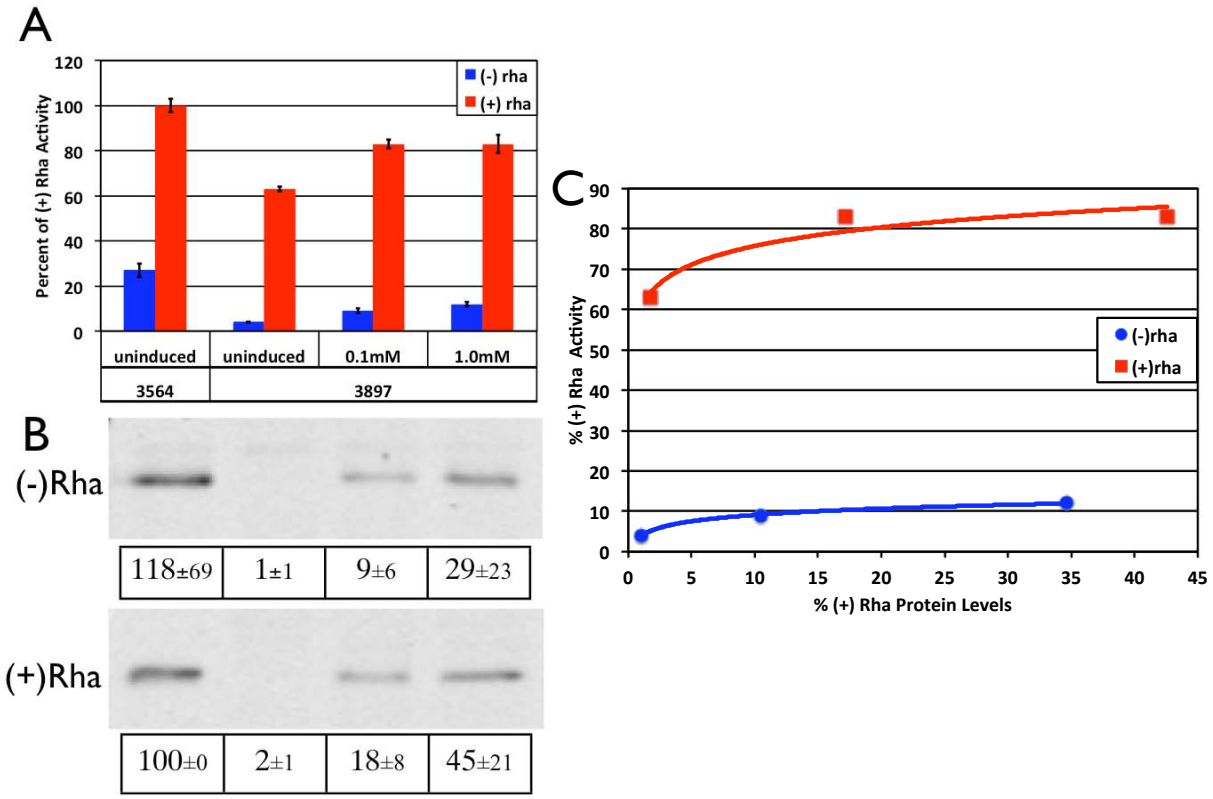


Figure 9. Effect of Protein Levels on RhaR Stability (-/+)Rhamnose. Activity and protein levels of RhaR were compared for a strain with a *malP*::*lacI*^{q} allele (SME3897) to those from a strain lacking the *malP*::*lacI*^{q} allele (SME3564). Expression of RhaR in SME3897 was induced with IPTG at the indicated concentrations. **A.** The averaged activity of RhaR from strain SME3564 was 1.9 Miller Units (-) Rha and 7.1 Miller Units (+) Rha. Values are the average of three independent assays with two replicates each and are shown as the percent of the (+) Rha value for SME3564. Error is shown as the standard deviation converted to percent of the Miller Unit values. **B.** Representative western blot and quantification of RhaR protein levels shown as the percent of the (+) Rha value for SME3564. RhaR values were normalized to DnaK levels and are the average from three replicates. Error is shown as the standard deviation converted to percent of the normalized protein values. **C.** RhaR protein levels vs. activation levels for SME3897 induced with IPTG (uninduced, 0.1 mM and 1.0 mM). All values are shown as the percent of the (+) Rha values for SME3564. Trend lines are the logarithmic regression of the (-) Rha or (+) Rha data points.

RhaR expression levels is approximately 6-fold (-)rhamnose and 1.6-fold (+)rhamnose, resulting in a maximal effect of a 3- to 4-fold activation in response to L-rhamnose for variants with decreased activity due only to stability. However, many of the variants with decreased basal activation (Down (-), Table 9) respond to L-rhamnose with a greater than 4-fold activation. This observation led us to hypothesize that the residues where these variants were isolated may be involved in interdomain contacts and are likely integral to the function of the (-)rhamnose state.

Perhaps the most interesting variants I isolated in the RhaR AS2 region are two variants at position H269 (H269S and H269K) with increased (-)rhamnose activity, but wild type activity (+)rhamnose (Table 9). Western blots indicated that neither variant had increased protein levels (-)rhamnose relative to wild-type RhaR, thus protein levels did not explain their increased transcription activation (Table 9). Although only two of the six variants identified at H269 exhibited increased (-)rhamnose activity (not the majority that would be expected for loss-of-residue-function variants), the remainder of the variants had decreased activity (-)rhamnose (Down (-)) and therefore could have stability defects that may have altered their phenotype. We propose that H269S and H269K are loss-of-residue-function variants and that H269 may be involved in inhibitory contacts in wild-type RhaR that reduce the activity of RhaR (-)rhamnose.

Site-Specific Random Mutagenesis of the C-terminal Ends of the RhaR Helix-Turn-Helix Motifs. In addition to the AraC and ToxT residues that align with RhaR AS2 (65-68), there are also ToxT residues near the C-terminal ends of the two helix-turn-helix motifs (HTH) that are in position to make interdomain contacts (HTH1), and/or enclose part of the effector-binding pocket in the ToxT NTD (HTH2) (79, 161). We refer to these

regions as C-terminal HTH1 (C-HTH1) and C-terminal HTH2 (C-HTH2), to distinguish them from the DNA-contacting residues that are generally nearer to the N-terminal ends of the DNA recognition helices. To determine whether any of the RhaR residues aligned with these ToxT residues have a role in allosteric L-rhamnose signaling, we again randomly mutagenized one codon at a time in each of these regions.

We mutagenized 5 positions in RhaR C-HTH2. Four of these positions align with ToxT residues that are positioned to contact the NTD or effector bound to the NTD (ToxT 258-261; PDB ID 3GBG). The fifth position (R295) is predicted to contact the NTD in I-TASSER models (159, 160) of RhaR, but not in ToxT – perhaps due to the shorter side chain at this position in ToxT (S257). We isolated 21 variants, and assayed them for transcription activation (-) and (+)rhamnose at the *rhaSR* promoter-*lacZ* fusion (Table 10). Similar to AS2, this region may also be important for RhaR function and/or stability, as three of the 21 C-HTH2 variants (14%) had decreased activity both (-) and (+)rhamnose. We measured protein levels for two of the three variants with this phenotype and found variant E296H had protein levels that were approximately wild-type or higher, while variant M299E had reduced protein levels that might explain its activity defects (Table 10). Similar to AS2, 60% (13 of 21) of the variants had decreased (-)rhamnose activity and wild-type activity (+)rhamnose. We measured protein levels for five of the thirteen variants with this phenotype and found that three variants had wild- type protein levels both (-) and (+)rhamnose (E296Y, T297S, G298L), while the other two variants had reduced (G298T) or elevated protein levels (M299V) (Table 10). Again, we hypothesize that residues where a majority of variants have decreased basal activation may be involved in interdomain contacts and are likely integral to the function of the (-)rhamnose state.

Three variants in the RhaR C-HTH2 region, all at position 295, had a phenotype that was not encountered in the Arm or AS2 regions: both decreased (-)rhamnose activation and increased (+)rhamnose activation (Table 10). Protein levels were assayed for two of these three variants and were not reduced (-)rhamnose for either variant (R295L and R295Q) (Table 10). For at least the R295Q variant, protein levels (+)rhamnose were not elevated, therefore the variant activity can not simply be explained by altered protein levels (Table 10). RhaR R295 aligns with a ToxT residue that has been implicated in DNA binding (S257) (79, 161) and with a MarA residue (N100) that is adjacent to a DNA-backbone contacting residue (MarA K99) (43, 163). Eighty eight percent of the ~200 AraC family protein sequences in the (former) AraC-XylS database (164) have one of the 6 residues most likely to contact the DNA backbone (R, K, S, T, N or Q) (165) at the position of RhaR R295, indicating that this position may be used to contact the DNA backbone in many homologs. We therefore propose that RhaR R295 may contact the DNA backbone in the (-)rhamnose state, explaining the finding that multiple substitutions resulted in decreased activity (-)rhamnose. Although the decreased activity of the three R295 variants (-)rhamnose could potentially be explained by stability defects, the evidence more strongly supports a role in DNA binding is more likely. The only variant at R295 with wild-type activity was the conservative substitution R295K, consistent with a role for this residue in contacting the DNA backbone. In the presence of rhamnose, the finding that multiple variants at RhaR R295 had increased activity indicated that they may be loss-of-residue-function variants, and thereby that the wild-type residue may decrease RhaR activity (+)rhamnose. We therefore propose that in the (+)rhamnose state, R295 either somewhat interferes with DNA binding or makes somewhat inhibitory contacts with the RhaR NTD, as

Table 10. Transcription Activation of *rhaS-lacZ* by RhaR C-HTH2 Variants[a]

Mutation	(-) Rha		(+) Rha		Category[c]
	% WT Activation[a]	% WT Protein Levels[b]	% WT Activation[a]	% WT Protein Levels[b]	
WT	100±14	100	100±6	100	
R295L	9±1	128±3	160±12	176±70	Down (-) Up (+)
R295I	10±1		169±22		Down (-) Up (+)
R295Q	25±2	238±76	262±41	94±16	Down (-) Up (+)
R295K	118±4	204±91	111±15	153±21	~WT
E296Y	14±7	113±33	4±1	117±75	Down
E296H	11±5	134±46	4±1	157±17	Down
E296L	14±2		91±11		Down (-)
E296Q	15±1		110±41		Down (-)
E296T	58±8	149±60	125±10	163±0.8	~WT
T297N	9±1		91±8		Down (-)
T297A	30±2	69±2	89±30	161±30	Down (-)
T297S	30±2	96±77	141±20	119±30	Down (-)
T297I	102±6	136±49	109±4	148±44	~WT
G298T	7±1	42±29	83±8	54±7	Down (-)
G298L	13±4	70±44	81±8	79±16	Down (-)
G298S	13±2		86±10		Down (-)
G298K	14±1		80±2		Down (-)
M299E	8±0.5	38±6	7±1	10±5	Down
M299T	30±3		96±9		Down (-)
M299L	50±4		111±22		Down (-)
M299V	51±2	231±142	121±6	158±0.7	Down (-)

a. The averaged activity of wild type RhaR was 2.3 Miller Units (-) Rha and 9.1 Miller Units (+) Rha. Values are the average of two independent assays with two replicates each and are shown as the percent of the corresponding (-) Rha or (+) Rha wild type values. Error is shown as the standard deviation converted to percent of the Miller Unit values.

b. Quantification of variant protein levels compared to wild-type RhaR. Proteins were transferred to nitrocellulose, probed with anti-RhaR and anti-DnaK antibodies and quantified using densitometry analysis. RhaR values were normalized to DnaK levels, are the average from two replicates and are shown as the percent of protein relative to the corresponding (-) Rha or (+) Rha wild type values. Error is shown as the standard deviation converted to percent of the normalized protein values. Western blots performed by Frances Mandelbaum.

c. Variants were categorized as follows: Down, reduced activity (-) and (+) Rha; Down (-), reduced activity (-) Rha, near wild-type (+) Rha; Down (-) Up (+), reduced activity (-) Rha, increased activity (+) Rha; ~WT, near wild-type activity (-) and (+) Rha.

further discussed below.

In the RhaR C-HTH1 region, I mutagenized five residues located at or near the position of ToxT residues that contact its NTD. Variants were assayed for transcription activation (-) and (+)rhamnose at (*rhaS-lacZ*)Δ85. Preliminary data indicated that, of 19 variants, only two variants (at position Q246) had interesting phenotypes (Table 11). A large number of variants showed decreased activity (-)rhamnose and near wild type activity (+)rhamnose; ten of the nineteen variants fell into this category. Therefore, we only performed final assays on the two interesting variants at position 246 (Q246S and Q246R) (Table 12). Variant Q246S had wild-type activity (-)rhamnose but 1.5-fold elevated activity in the presence of effector while Q246R had a 3-fold decrease in activity (-)rhamnose and a large, 6-fold increase in activity (+)rhamnose. Protein levels were approximately wild type for both variants (Table 12), indicating that altered protein levels do not explain the phenotypes. Although the phenotype of Q246R is similar to the RhaR R295 variants, the aligned position (E51) in the MarA structure (43) is not in a position to contact the DNA. However, the aligned position in ToxT (R209) is in position to contact the ToxT NTD. The RhaR Q246R variant may have gained an interaction with the NTD that enhances the activity of the RhaR (+)rhamnose state to the detriment of the (-)rhamnose state. Given our prediction that this is a gain-of-residue-function variant that might contact the NTD (+)rhamnose, we feel it is best not to use the ToxT structure (79) to make predictions regarding any potential residues that may be contacted by Q246R. This is because the ToxT structure represents the non-inducing state, and we have no information about the relative orientation of the two domains in the (+)rhamnose state. The differential phenotypes of Q246R (-) and (+)rhamnose suggests that this region of RhaR

may undergo a conformational change in response to L-rhamnose.

Site-Specific Random Mutagenesis of RhaR AS1. We continued our approach of random mutagenesis to determine if residues in the allosteric site in subdomain 1 (AS1) (RhaR residues 208-218) also are involved in making L-rhamnose-dependent interdomain contacts. Similar to residues in the AS2, C-HTH1, and C-HTH2 regions, residues in AS1 align with residues in ToxT that are positioned to make interdomain contacts with the ToxT NTD (79). At AS1, site-directed random mutagenesis was carried out at three positions (D209, T213, and A217) and a total of eight variants were identified (Table 11). Variants were assayed for transcription activation (-) and (+)rhamnose at (*rhaS-lacZ*)Δ85. Within AS1, these three residues were chosen for mutagenesis based on models made with I-TASSER protein structure prediction program (159, 160) and modeling using the MarA structure (43) which predicted that these three residues were likely surface exposed and not buried in the structure. Variants at D209 had decreased activity (-)rhamnose while retaining wild type or higher activity in the presence of the effector, similar to the variants found in the Arm, AS2, and the C-HTH2. Mutagenesis of T213 yielded only two variants (each identified multiple times from two separate rounds of cloning). Substitution at position 213 with leucine resulted in decreased activity (-) but not (+)rhamnose, while substitution with alanine gave wild type activity both with and without the effector. The repeated isolation of the same two substitutions at position 213 suggests the possibility that other substitutions at this position may be toxic to the RhaR protein, which may indicate a role for this residue in protein folding and/or stability considering only conservative substitutions were tolerated. Two of the three of the variants at A217 had near wild-type activity both (-) and (+)rhamnose, indicating that this residue does not have a significant

Table 11. Transcription Activation of *rhaS-lacZ* by RhaR C-HTH1 Variants[a]

Mutation	% WT Activation		Category[b]
	(-) Rha	(+) Rha	
WT	100	100	
F243I	14	33	Down
F243L	16	59	Down (-)
F243M	17	107	Down (-)
Q245G	43	81	Down (-)
Q245V	59	82	Down (-)
Q245C	66	74	Down (-)
Q245A	98	84	~WT
Q246L	51	113	Down (-)
Q246R	25	591	Down (-) Up (+)
Q246S	70	154	Up (+)
Q246W	76	117	~WT
T247G	13	3	Down
T247W	17	5	Down
T247A	15	76	Down (-)
T247I	22	72	Down (-)
Q257I	19	31	Down
Q257L	23	90	Down (-)
Q257R	25	119	Down (-)
Q257V	99	88	~WT

a. Variants were assayed in groups with a wild type RhaR activity range of 1.1-1.5 Miller Units (-) Rha and 7.3-8.3 Miller Units (+) Rha. Values are shown as the percent of the corresponding (-) Rha or (+) Rha wild type values. These values are preliminary data and are the result of a single assay with two replicates.

b. Variants were categorized as follows: Down, reduced activity (-) and (+) Rha; Down (-), reduced activity (-) Rha, near wild-type (+) Rha; Down (-) Up (+), reduced activity (-) Rha, increased activity (+) Rha; ~WT, near wild-type activity (-) and (+) Rha; Up (+), wild-type activity (-) Rha, increased activity (+) Rha.

Table 12. Transcription Activation of *rhaS-lacZ* by RhaR Variants Q246R and Q246S[a]

Mutation	(-) Rha		(+) Rha		Category[c]
	% WT Activation	% WT Protein[b]	% WT Activation	% WT Protein[b]	
WT	100±18	100	100±24	100	
Q246R	36±9	159±29	658±15	108±25	Down (-) Up (+)
Q246S	77±7	93±15	149±10	61±17	Up (+)

a. The averaged activity of wild type RhaR was 1.1 Miller Units (-) Rha and 5.4 Miller Units (+) Rha. Values are the average of two independent assays with two replicates each and are shown as the percent of the corresponding (-) Rha or (+) Rha wild type values. Error is shown as the standard deviation converted to percent of the Miller Unit values.

b. Quantification of variant protein levels compared to wild-type RhaR. Proteins were transferred to nitrocellulose, probed with anti-RhaR and anti-DnaK antibodies and quantified using densitometry analysis. RhaR values were normalized to DnaK levels, are the average from two replicates and are shown as the percent of protein relative to the corresponding (-) Rha or (+) Rha wild type values. Error is shown as the standard deviation converted to percent of the normalized protein values.

c. Variants were categorized as follows: Down (-) Up (+), reduced activity (-) Rha, increased activity (+) Rha; Up (+), wild-type activity (-) Rha, increased activity (+) Rha.

Table 13. Transcription Activation of *rhaS-lacZ* by RhaR AS1 Variants[a]

Mutation	% WT Activation		Category[b]
	(-) Rha	(+) Rha	
WT	100	100	
D209A	12	110	Down (-)
D209G	16	183	Down (-) Up (+)
D209E	27	128	Down (-)
T213L*	18	111	Down (-)
T213A*	141	86	~WT
A217P	14	114	Down (-)
A217V	81	94	~WT
A217G*	156	119	~WT

a. Variants were assayed in groups with a wild type RhaR activity range of 1.5-2.5 Miller Units (-) Rha and 7.5-10.6 Miller Units (+) Rha. Values are shown as the percent of the corresponding (-) Rha or (+) Rha wild type values. These values are preliminary data and are the result of a single assay with two replicates.

b. Variants were categorized as follows: Down (-), reduced activity (-) Rha, near wild-type (+) Rha; ~WT, near wild-type activity (-) and (+) Rha. *Denotes that values are the average of two variants with the same substitution assayed simultaneously

role in allosteric signaling in RhaR.

Screen for Second Site Suppressors. From mutagenesis of the RhaR Arm and the mutagenesis I performed in the DBD, we identified positions in the Arm and AS2 with predicted roles in interdomain allosteric L-rhamnose signaling. We hypothesized that these RhaR Arm and AS2 residues might make direct contact with each other based on studies with AraC and ToxT, in addition to computational predictions of the RhaR structure where the RhaR Arm is positioned to contact AS2 (Figure 10). Many studies provide evidence that the AraC Arm contacts a region in the DBD that aligns with AS2 (25, 64-68), and structures of the AraC NTD in the absence and presence of arabinose show that the Arm is positioned over or adjacent to the effector binding pocket (70, 71). In addition, although ToxT does not have the equivalent to an Arm, the structure of full length ToxT shows that the effector-binding pocket, where the AraC Arm is located, is in close proximity with the ToxT region that aligns with RhaR AS2 (79).

I tested whether it was possible to isolate second site suppressors of the variants at positions L35 in the RhaR Arm and H269 in AS2 that were implicated in inhibition of transcriptional activation by RhaR (-)rhamnose (Table 10 and Koppolu and Egan, manuscript in preparation). If residues L35 and H269 interact in the wild type protein, then we should be able to identify second site suppressor mutations that restore this contact in mutant alleles where this contact has been disrupted. Thus, I combined: L35K with a randomly mutagenized codon at position 269 (H269X); a randomly mutagenized codon at position 35 (L35X) with H269S; and L35X with H269K. I tested L35K since this transcription activation (-) and (+)rhamnose at the *rhaSR* promoter fused with *lacZ* and variant had the largest impact on (-)rhamnose activation. Candidates were assayed for

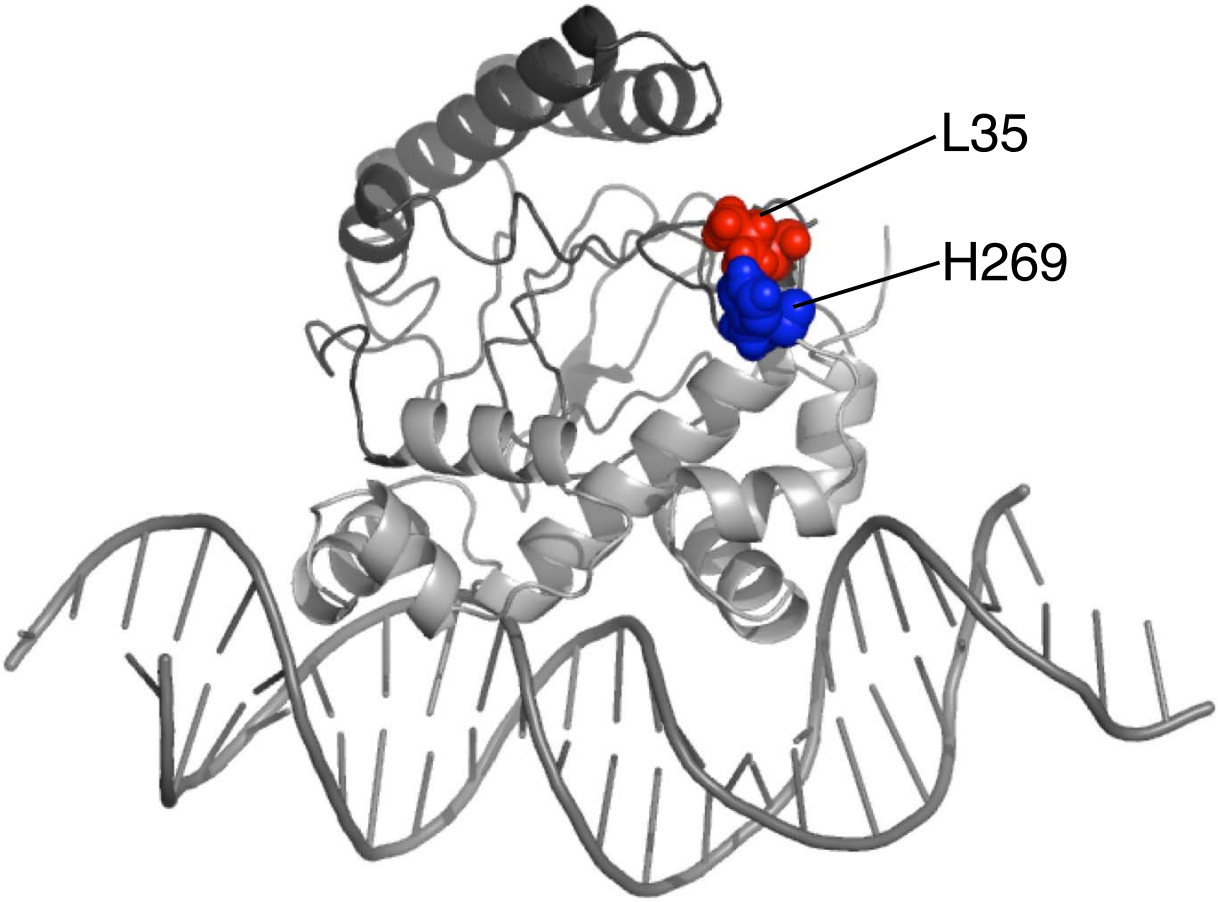


Figure 10. Contact Between L35 and H269 in the Predicted RhaR Structure. A structural model of the RhaR protein from the I-TASSER structural prediction program predicting contact between residues L35 in the RhaR Arm and residue H269 in the AS2 region of the DBD. The RhaR NTD is colored dark gray and the DBD is colored light gray. Residues L35 (red) and H269 (blue) are shown as space filling spheres. The predicted RhaR structure was modeled onto DNA from the MarA co-crystal (43).

those that suppressed the elevated (-)rhamnose activation (in other words, reverted to wild-type (-)rhamnose activation) were sequenced to identify the mutation.

In my screen for suppressors of L35K (L35K combined with H269X), I identified H269V as a potential second site suppressor (Table 14). The H269V substitution suppressed elevated basal activation by L35K back to the wild-type level. In this case, true second site suppressors would be defined as not decreasing activity in the absence of the first substitution, thus we constructed a variant with H269V in the absence of L35K. We found that H269V had approximately wild-type activity (-)rhamnose, providing support that it is a true second site suppressor. We hypothesize that the L35K-H269V variant restored interdomain contacts, made between the residues at positions 35 and 269 in wild type RhaR, but lost in L35K, that contribute to the inhibition of RhaR activity (-)rhamnose.

Unlike position L35, where multiple substitutions resulted in increased activity (-)rhamnose, (and thus were considered loss-of-residue function mutations), only one substitution at position L37 had increased activity (and thus was considered a gain-of-function mutation) (Koppolu and Egan, manuscript in preparation). A position with a gain-of-function mutation that increases activity (-)rhamnose is not predicted to be involved in an interdomain contact in wild-type RhaR. As a control, I screened for second site suppressors of the gain-of-residue function variant L37R in the Arm. I combined: L37R with randomly mutagenized codons at position 269 (H269X); L37X with H269S; and L37X with H269K. Despite screening comparable numbers of colonies as the screen for second-site suppressors of L35K, I did not identify candidates at positions 37 or 269 that suppressed the elevated (-)rhamnose activation of variants at the other position, suggesting that L37 and H269 likely do not make contacts that contribute to inhibition of

Table 14. Transcription Activation of *rhaS-lacZ* by RhaR Second-Site Suppressor Variants[a]

Mutation	% WT Activation	
	(-) Rha	(+) Rha
WT	100±4	100±10
L35K	190±4	117±9
H269V	96±4	117±10
L35K-H269V	115±13	122±2

a. The averaged activity of wild type RhaR was 1.7 Miller Units (-) Rha and 7.8 Miller Units (+) Rha. Values are the average of two independent assays with two replicates each and are shown as the percent of the corresponding (-) Rha or (+) Rha wild type values. Error is shown as the standard deviation converted to percent of the Miller Unit values.

b. Quantification of variant protein levels compared to wild-type RhaR. Proteins were transferred to nitrocellulose, probed with anti-RhaR and anti-DnaK antibodies and quantified using densitometry analysis. RhaR values were normalized to DnaK levels, are the average from two replicates and are shown as the percent of protein relative to the corresponding (-) Rha or (+) Rha wild type values. Error is shown as the standard deviation converted to percent of the normalized protein values. Western blots were performed by Frances Mandelbaum.

RhaR activity (-)rhamnose. The inability to identify second site suppressors of L37R supports the hypothesis that the H269V is a true second site suppressor of L35K, and that L35 and H269 may make contacts (-)rhamnose that contribute to maintaining the low basal activity of RhaR (-)rhamnose.

Screen for Epistatic Contacts. As an alternative method to test our hypothesis that RhaR Arm residues interact with residues in AS2 to inhibit transcription (-)rhamnose, I also combined variants with increased basal activation at each position to look for epistatic effects. We expected that, if these residues do interact to inhibit transcription activation, combining two loss-of-function variants would not further increase basal activation, as both residues are defective for the same contact or function. However, if the variant with combined substitutions does have additive effects (further increasing basal activation) the two residues are likely involved in different contacts or functions. When the L35K variant was combined with H269S and H269K, the effects were non-additive, as we would expect if these residues interact (Table 15). I also constructed a series of controls using variants with increased basal activity isolated in regions of the protein that we did not expect to interact with the DBD, based on structural models and comparison to the ToxT crystal structure (79) (Figure 11).

As controls, I combined W75R with both H269K and H269S and D197G with H269K (Table 15). The W75R variant has increased basal activity and may have a role in allosteric signaling (144). Residue W75 maps to the β-barrel sugar-binding subdomain of the NTD on the AraC-NTD structure with arabinose bound (71, 144) and based on the ToxT structure (79), should not be in close enough proximity to contact H269. The RhaR D197G variant also has slightly elevated basal activity and is predicted to lie at the very end of the

Table 15. Transcription activation of *rhaS-lacZ* by RhaR Arm-AS2 Epistasis Variants[a]

Mutation	(-) Rha		(+) Rha	
	% WT Activation	% WT Protein[b]	% WT Activation	% WT Protein[b]
WT	100±16	100	100±12	100
L37R	151±22	32±4	61±2	40±4
H269K	162±37	65±3	96±5	95±28
H269S	178±28	86±12	78±3	94±35
L37R-H269K	145±7	21±2	72±5	8±6
L37R-H269S	198±12	26±5	73±4	18±8
W75R	269±17	52±7	168±12	85±5
W75R-H269K	312±27	35±1	166±1	89±32
W75R-H269S	313±29	42±7	170±1	96±17
D197G	139±21	64±3	80±7	67±20
D197G-H269K	161±16	27±1	85±4	59±18

a. The averaged activity of wild type RhaR was 1.8 Miller Units (-) Rha and 9.2 Miller Units (+) Rha. Values are the average of two independent assays with two replicates each and are shown as the percent of the corresponding (-) Rha or (+) Rha wild type values. Error is shown as the standard deviation converted to percent of the Miller Unit values.

b. Quantification of variant protein levels compared to wild-type RhaR. Proteins were transferred to nitrocellulose, probed with anti-RhaR and anti-DnaK antibodies and quantified using densitometry analysis. RhaR values were normalized to DnaK levels, are the average from two replicates and are shown as the percent of protein relative to the corresponding (-) Rha or (+) Rha wild type values. Error is shown as the standard deviation converted to percent of the normalized protein values. Western blots were performed by Frances Mandelbaum.

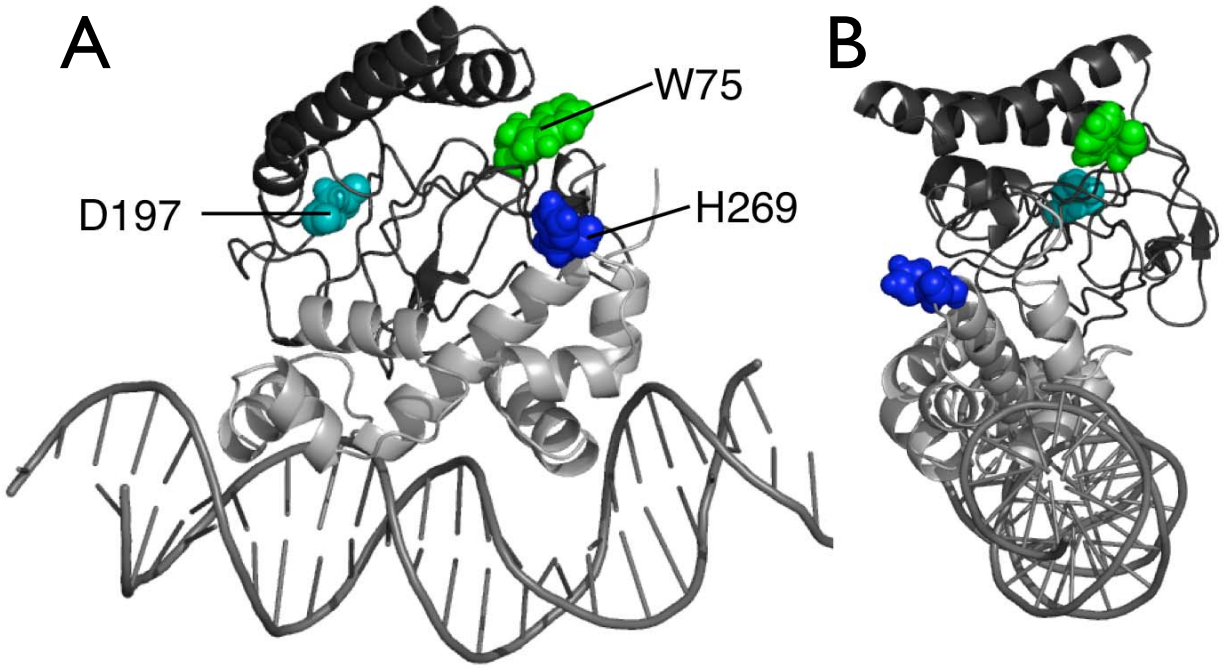


Figure 11. Positioning of residues W75 and D197 relative to H269 in the RhaR structural prediction model. A structural model of the RhaR protein from the I-TASSER structural prediction program. Residues W75 and D197 (in the RhaR NTD) are not predicted to be in a position to contact residue H269 (in the RhaR DBD). The RhaR NTD is colored dark gray and the DBD is colored light gray. Residues W75 (green), D197 (cyan) and H269 (blue) are shown as space filling spheres. The predicted RhaR structure was modeled onto DNA from the MarA co-crystal (43). **A.** "Front" view. The HTH motifs project to the forefront. **B.** "Side" view. Image rotated 90° relative to the "Front" view (A.)

RhaR NTD or within the beginning of the flexible linker connecting the RhaR-NTD to the DBD (71, 144, 166). Based on the predicted location of D197, we hypothesized that it was unlikely that this residue interacts with H269. However, both sets of controls gave non-additive effects indicating that these variants increase basal activity through the same global mechanism. Therefore, it is not possible to draw any conclusions regarding the interaction of the RhaR Arm with the DBD from these experiments.

To further probe the interaction of the Arm with the DBD, I also combined variants D41A and D41S with T279A and T279S, respectively. Although structural modeling of the RhaR protein did not indicate interaction of the Arm with T279 (Figure 12), the presumed flexibility of this region combined with multiple positions of the Arm within the various structural models predicted by I-TASSER, led us to hypothesize that perhaps residue D41 in the Arm contacts residue T279 to modulate σ^{70} contacts for transcription activation. T279 is located between two known σ^{70} contacting residues (37, 155) in a structural model indicating that the side chain at this position may cause steric interference with σ^{70} contacts (-)rhamnose (144) (Figure 12).

In the Arm, D41A had elevated activity both (-) and (+)rhamnose while D41S had only elevated activity in the absence of effector (Koppolu and Egan, manuscript in preparation). In the DBD, T279 is the only position other than H269 where variants with increased activity (-)rhamnose have been isolated. Both T279A and T279S were isolated from a previous screen searching for variants with elevated basal activation to identify residues involved in allosteric signaling (144). We noted that both positions had alanine and serine substitutions that resulted in elevated basal activation. Given the strong contact propensity between two alanines or two serines (1.2 and 1.6, respectively) (167), we chose

Table 16. Transcription Activation of *rhaS-lacZ* by RhaR Arm-T279 Epistasis Variants

Mutation	% WT Activation	
	(-) Rha	(+) Rha
WT	100	100
D41A	2.5	8.5
D41S	2.8	9
T279A	16	64.5
T279S	3.2	98.2
D41A-T279A	20.7	56
D41S-T279S	15.4	43.5

The averaged activity of wild type RhaR was 1.6 Miller Units (-) Rha and 5.7 Miller Units (+) Rha.. Values are shown as the percent of the corresponding (-) Rha or (+) Rha wild type values. These values are preliminary data and are the result of a single assay with two replicates.

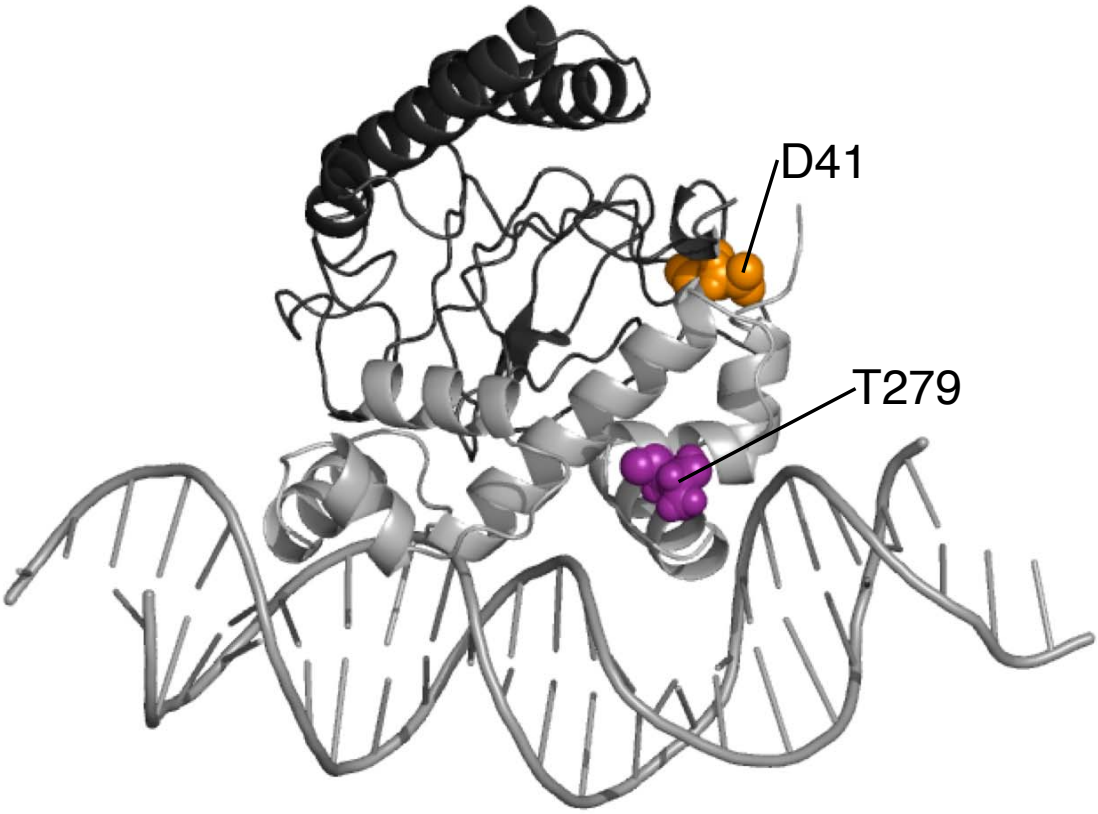


Figure 12. Positioning of Residues D41 and T279 in the RhaR Structural Prediction Model. A structural model of the RhaR protein from the I-TASSER structural prediction program. Although the model does not predict interaction of residues D41 (in the RhaR Arm) and T279 (in the RhaR DBD), we hypothesized that (-)rhamnose, the Arm could be positioned such that it could contact T279 to inhibit transcription activation through modulation of contacts with RNAP σ^{70}. The RhaR NTD is colored dark gray and the DBD is colored light gray. Residues D41 (orange), and T279 (purple) are shown as space filling spheres. The predicted RhaR structure was modeled onto DNA from the MarA co-crystal (43).

to combine variants with alanine or serine substitutions at D41 and T279. Both D41A-T279A and D41S-T279S had additive effects when compared to each variant alone (Table 16). We therefore conclude that D41 likely does not directly interact with T279 to modulate σ^{70} contacts.

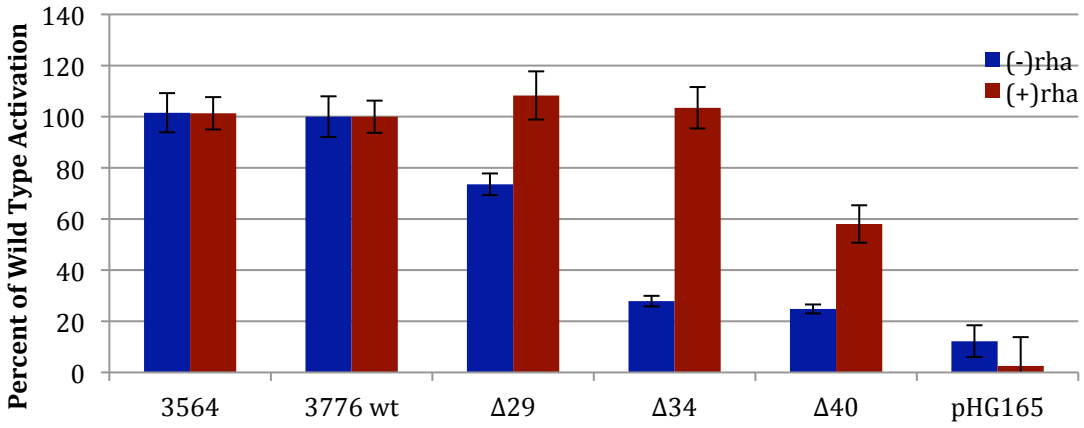
Determination of the Role of the N-terminal Extension of RhaR. A series of nested N-terminal truncations of RhaR were previously constructed, deleting from six to fifty-two amino acids from the N-terminus of the protein to determine if RhaR uses a mechanism of allosteric signaling similar to AraC (144). Compared to AraC, RhaR has an additional twenty-nine residues at its N-terminus and an additional thirty-four residues compared to RhaS. Therefore, a large range of deletions were made to determine which residues, if any, were acting as the N-terminal Arm. Unlike AraC (64, 71), N-terminal truncation of RhaR did not result in constitutive activation in the absence of L-rhamnose. Deletion of the first 40 amino acids resulted in reduced activation levels both in the presence and absence of L-rhamnose, although protein levels were lower than those of the wild type (144).

In the previous study of the N-terminal deletions of RhaR, variants were expressed in a high copy number plasmid, pGEM (144). As a result, there was concern that increased protein levels may mask potential defects of the truncated variants, which might be detectible at lower protein levels. I therefore, re-constructed three of the deletions (Δ2-29, Δ2-34, Δ2-40) and cloned them into pHG165, which has a lower copy number than the pGEM plasmid previously used. I then tested the level to which these variants activated transcription of two transcriptional fusions, Φ(*rhaS-lacZ*)Δ85 and Φ(*rhaS-lacZ*)Δ128 (both (-) and (+)rhamnose) to determine what effect, if any, the extension has on CRP

co-activation. Fold activations were similar both (-) and (+)rhamnose for both transcriptional fusions (Figure 13) and were similar to previous findings (144). None of the deletions resulted in constitutive activity (-)rhamnose. Deletion of the first thirty-four amino acids had at most, a slight effect on activation (+)rhamnose, but further deletion of the first forty amino acids resulted in decreased activity (+)rhamnose. Western blot analyses were performed on this set of variants, again in both transcriptional fusions. Protein levels were greatly reduced (-)rhamnose for all three deletions (Figure 14). Furthermore, protein levels were reduced (+)rhamnose with levels decreasing upon further truncation. From these results, we concluded that the first thirty-four amino acids are not necessary for stimulation (+)rhamnose, but protein instability upon further deletion of RhaR hindered our ability to draw conclusions regarding the role for the N-terminal Arm (residues 35-41) of RhaR. This conclusion is consistent with the recent finding that RhaR residues 1-31 are naturally cleaved from the protein in *E. coli* and that only the cleaved species is likely active (Li and Egan, manuscript in preparation).

CRP Co-Activation to Determine if RhaR bends DNA Differentially (+/-)rhamnose. Maximal expression of *rhaSR* requires both RhaR and CRP, in addition to the α-CTD of RNA polymerase (31, 95). Co-activation with CRP at *rhaSR* results in an approximate three-fold increase (-)rhamnose, while (+)rhamnose CRP co-activation is approximately 100-fold (31). Activation by CRP at *rhaSR* requires both RhaR and α-CTD in addition to CRP (31, 95). The position of the *rhaSR* CRP binding site at -111 (relative to the transcription start site) is too far upstream to contact α-CTD without additional bending (31, 95, 168). RhaR has been shown to bend DNA by 160°, which likely allows CRP to interact with α-CTD (25). Due to the large difference in CRP co-activation in the presence

A



B

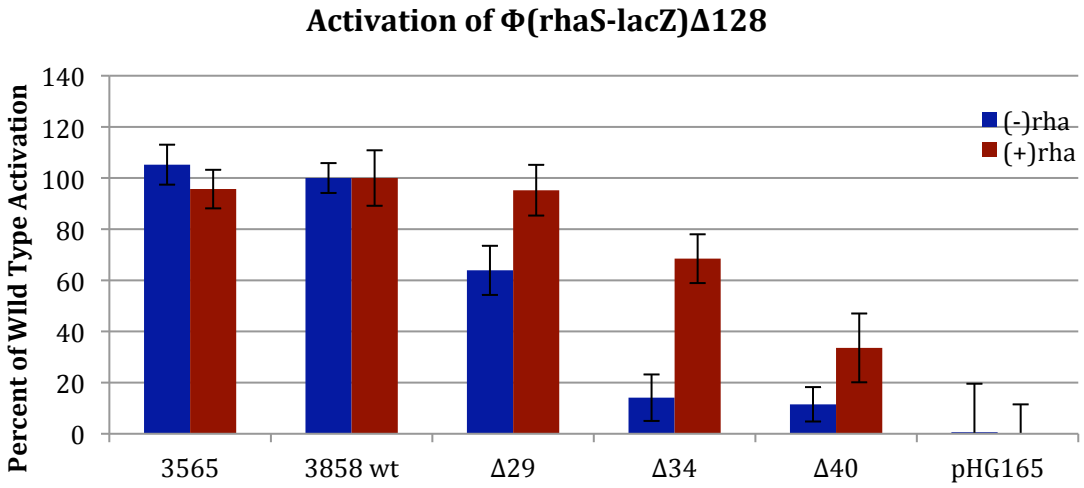


Figure 13. ***In Vivo*** **Transcription activation by RhaR N-terminal Deletion Variants.**

The averaged activity of wild type RhaR was 1.8 Miller Units (-) Rha and 9.2 Miller Units (+) Rha. Values are the average of two independent assays with two replicates each and are shown as the percent of the corresponding (-) Rha or (+) Rha wild type values. Error (SEM) was no greater than 19%. **A**: Activation of Φ(*rhaS-lacZ*)Δ85 in strain SME3160. The averaged activity of wild type RhaR (strain SME3776) in Miller units was 1.5 (-) Rha and 6.9 (+) Rha. **B**: Activation of Φ(*rhaS-lacZ*)Δ128 in strain SME2525. The averaged activity of wild type RhaR (Strain SME3858) in Miller units was 132 (-) Rha and 648 (+) Rha.

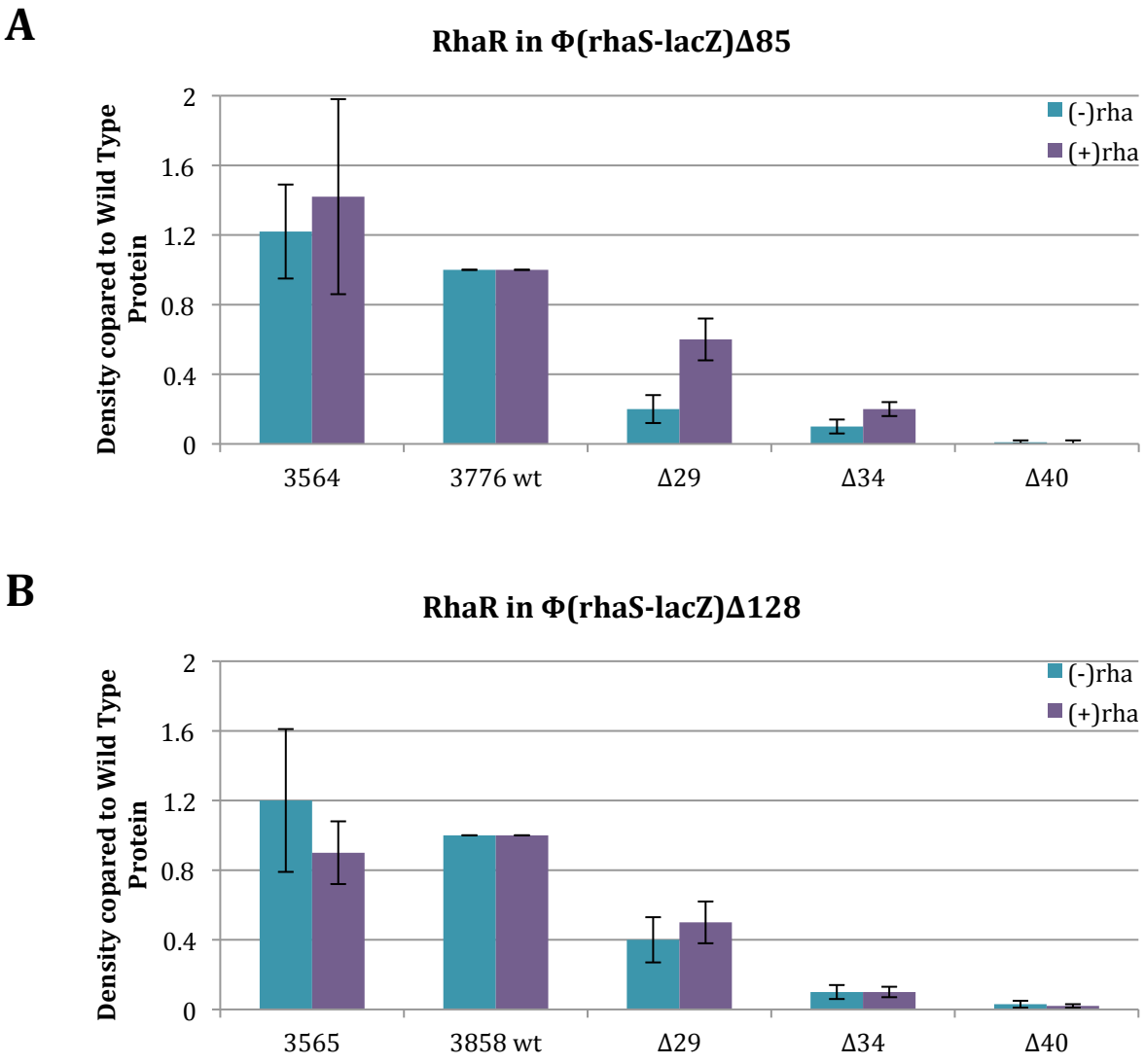


Figure 14. Quantification of RhaR N-terminal Deletion Variant Protein Levels.

Quantification of variant protein levels compared to wild-type RhaR. Proteins were transferred to nitrocellulose, probed with anti-RhaR and anti-DnaK antibodies and quantified using densitometry analysis. RhaR values were normalized to DnaK levels, are the average from three replicates and are shown as the band density relative to the corresponding (-) Rha or (+) Rha wild type bands. Error (SEM) was no greater than 19%. A: Protein levels in strain SME3160. B. Protein levels in strain SME2525.

and absence of L-rhamnose, we hypothesized that perhaps the extent to which RhaR bends DNA differs between these two states. We further hypothesized that the phenotype of variants with decreased basal activity but wild-type activation (+)rhamnose may be explained by defects in DNA bending in the (-)rhamnose state compared to wild type RhaR. (This hypothesis was developed prior to performing the experiments in Figure 9). To test this hypothesis, I first sought to determine if the amount of CRP co-activation is dependent upon RhaR protein levels. The differences in CRP activation previously reported were measured when RhaR was expressed chromosomally (31), however a difference in CRP co-activation (-) and (+)rhamnose was not observed when RhaR was expressed from a plasmid. To further investigate, I measured RhaR activation when RhaR was genomic or expressed from a plasmid in strains with (Δ128) and without (Δ92) CRP binding sites (Table 17) and determined the fold activation by CRP (Table 18). Preliminary results confirmed that moderate overexpression from the plasmid eliminated the difference in CRP co-activation (-) and (+)rhamnose (Table 18). Although the fold CRP activation (+)rhamnose with genomic RhaR was smaller than that previously reported, the decreased fold activation is likely due to lower activation of the Δ128 *rhaS-lacZ* fusion (+)rhamnose than previously measured (31). Activation levels (-)rhamnose are consistent with previous findings (31). Our preliminary findings indicated that differences in the CRP activation previously reported are sensitive to the RhaR protein levels and likely are not due to differential bending of the DNA in the effector bound and unbound states.

Electrophoretic Mobility Shift Assays with GB1basic-RhaR and SE-1. A small molecule inhibitor, SE-1 (previously referred to as OSSL_051168), was identified using whole-cell assays in a high throughput screen to identify inhibitors of the RhaS protein

Table 17. CRP Co-Activation of *rhaS-lacZ*

Promoter fusion	*rhaR*	β-Galactosidase sp act	
		(-)Rha	(+)Rha
(*rhaS-lacZ*)Δ85	Δ*rhaSR*	0.2	0.5
	chromosomal	0.2	4.7
	pHG165	N/D	N/D
	pHG165/*rhaR*	2.1	8
(*rhaS-lacZ*)Δ92	Δ*rhaSR*	0.1	0.09
	chromosomal	0.1	3.7
	pHG165	0.1	0.1
	pHG165/*rhaR*	1.2	6.1
(*rhaS-lacZ*)Δ128	ΔrhaSR	1.1	1.1
	chromosomal	0.3	71
	pHG165	1.3	1.2
	pHG165/rhaR	200	757

β-Galactosidase specific activity was assayed from a single-copy Φ(*rhaS-lacZ*) fusion with (Δ128) or without (Δ85 and Δ92) a CRP binding site, in a strain with either chromosomal *rhaR* and *recA*::*cat*, or Δ(*rhaSR*)::*kan recA*::*cat* and wt RhaR expressed from plasmid pHG165. Cultures were grown with [(+) Rha] or without [(-) Rha] L–rhamnose. These values are preliminary data and are the result of a single assay with two replicates.

Table 18. Fold Activation by CRP.

rhaR	Fold CRP Activation	
	(-)Rha	(+)Rha
ΔrhaSR	11	12
chromosomal	3	19
pHG165	13	12
pHG165/rhaR	167	124

The fold activation by CRP alone was determined by dividing the Miller unit value of the strain with CRP co-activation (Δ128) by the strain without CRP co-activation (Δ92) both in the presence and absence of L–rhamnose.

(148). SE-1 blocked RhaS-GB1^{201b} DNA binding in Electrophoretic Mobility Shift Assays (EMSA) but did not block DNA binding by the non-AraC family proteins LacI and CRP, indicating that the inhibitor likely blocks the function of the conserved AraC family DNA binding domain (DBD) (148). To determine if *in vitro* DNA binding by RhaR was blocked by SE-1, I purified GB1^{basic}-RhaR (GB1^{b}-RhaR) and performed Electrophoretic Mobility Shift Assays (EMSAs) to measure *in vitro* DNA binding in the presence of SE-1 (148).

The purification method for GB1^{b}-RhaR was developed by Dr. James Deng and was performed as previously described (148). GB1^{b}-RhaR was expressed from the vector pSE290/*rhaR* in ArcticExpress (DE3) cells (Agilent Technologies, Santa Clara, CA) in the presence of L-rhamnose. The purified protein was soluble and active for DNA binding and represented approximately 20% of the final purified protein; the cold-adapted chaperonin from ArcticExpress, Cpn60, co-purified with GB1^{b}-RhaR and represented approximately 50% of the final purified protein (Figure 15).

After purifying GB1^{b}-RhaR, I proceeded with testing *in vitro* DNA binding by RhaR in the presence of SE-1 (Figure 15) (148). Double stranded DNA encoding the RhaR binding sites at *rhaSR* (including binding sites for both monomers of the RhaR dimer) was incubated with GB1^{b}-RhaR in the absence or presence of increasing concentrations of the inhibitor (Figure 16-A). RhaR was fully inhibited by SE-1 in a dose-dependent manner with an IC_{50} of approximately 140μM (Figure 16-B) (148). Given this result, we concluded that SE-1 likely inhibits RhaR activity by blocking binding of the DBD to DNA. Given that SE-1 also blocks RhaS activity, but did not block activity of unrelated proteins, it is likely SE-1 inhibits DNA binding for additional AraC family proteins (148).

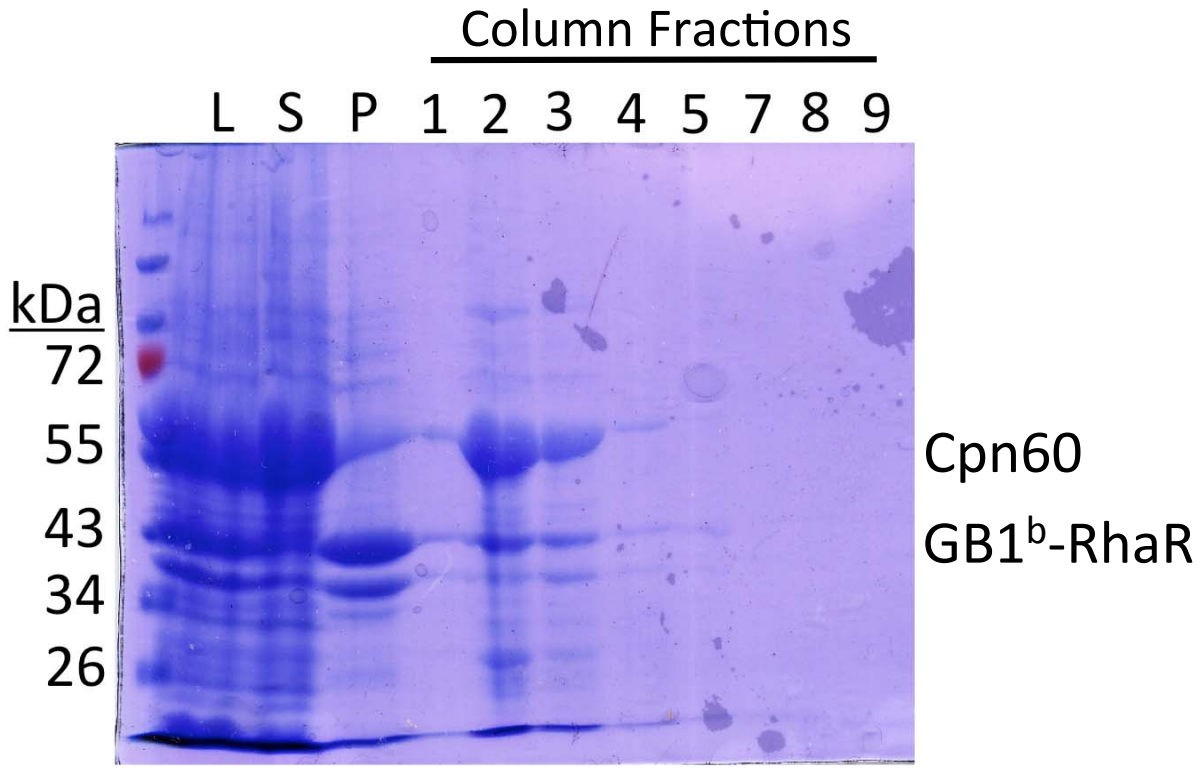
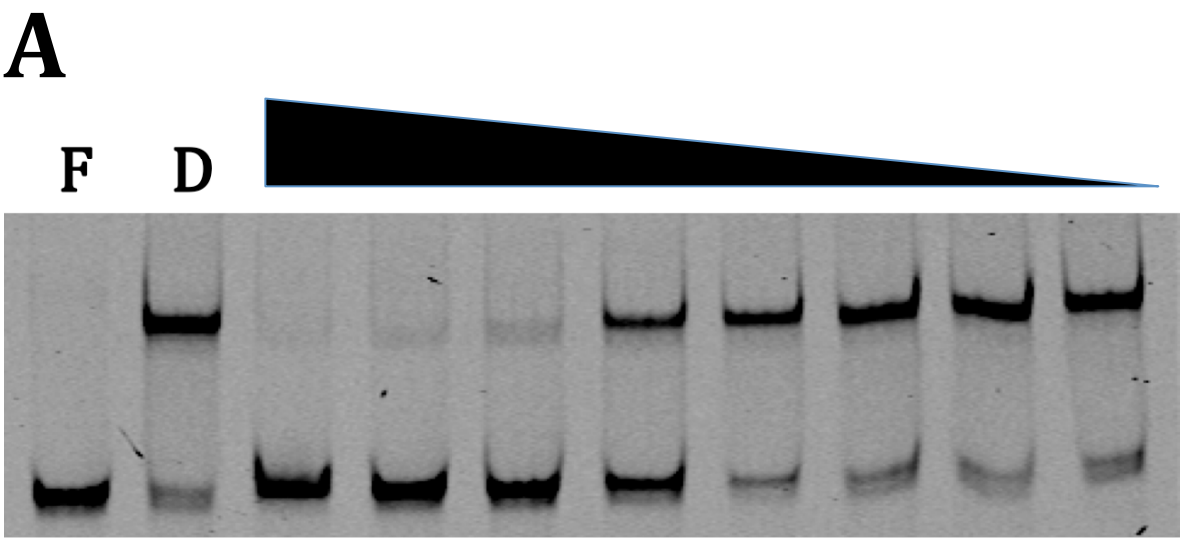


Figure 15. Expression and Purification of GB1^{b}-RhaR. Column fractions from purification of GB1^{b}-RhaR were electrophoresed on a 12% SDS-polyacrylamide gel and stained with Coomassie Blue. Fractions were eluted with 20mM Tris, 500mM NaCl, 250mM Imidazole, pH 7.9. L, sonicated lysate; S, supernatant fraction of sonicated lysate; P, insoluble pellet fraction of sonicated lysate. Column fractions are numbered according to the order in which they were eluted. GB1^{b}-RhaR was expressed from plasmid pSE290 in strain ArcticExpress (DE3). Gel is representative from one of two independent experiments.



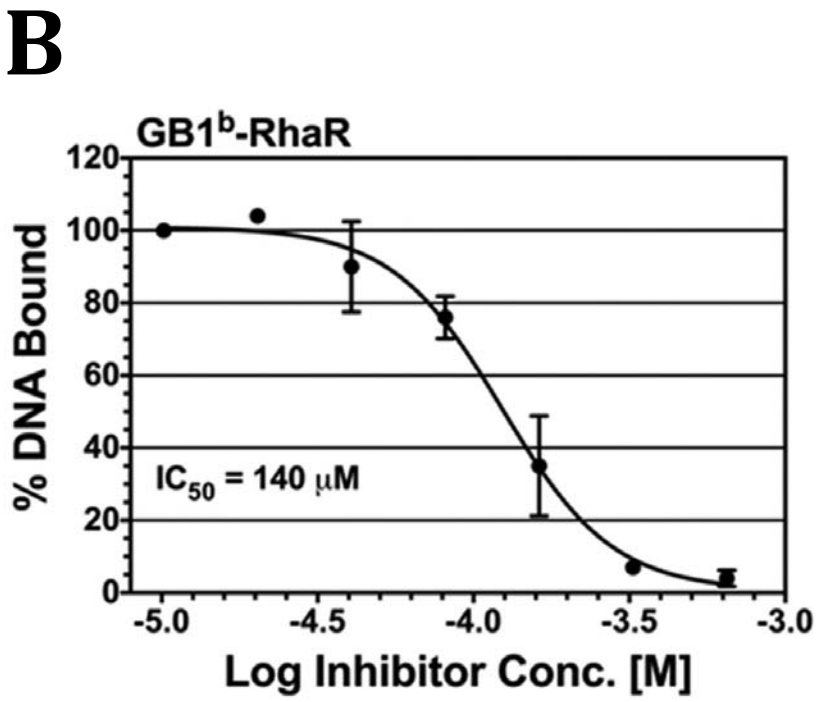


Figure 16. Electrophoretic Mobility Shift Assay to Measure Inhibition of *In Vitro* DNA Binding by RhaR in the Presence of SE-1. The DNA fragment was generated by annealing oligos 3056 (IR700 labeled) and 3287. GB1^{basic}-RhaR was used at 2.2μM with DNA at 1.3nM. Results are the average of three independent experiments. **A.** F, free DNA. D, DMSO only control. Inhibitor concentrations ranged from 1,300 to 10μM with two-fold serial dilutions. **B.** The DNA bound at the lowest concentration of inhibitor was set to 100% and binding in all other cases is relative to that value. Quantification and graphing done by Veerendra Koppolu. From (148).

Chapter IV: Mechanism of Transcription Activation in VirF

Activation by VirF-DBD Versus Full-Length VirF. The VirF protein of *Shigella* is required for transcription activation of genes necessary for successful invasion and infection by *Shigella* (169, 170). Regulation of virulence genes by VirF occurs in response to temperature, with maximal activation occurring at 37°C (106, 171). Previous reports have suggested that the N-terminal residues of VirF are likely responsible for protein dimerization and response to environmental stimuli, such as temperature (130, 172). To determine whether the VirF-DBD alone was sufficient to activate transcription, I constructed a variant encoding only the conserved AraC family DBD (Δ2-160) in the plasmid pHG165. This plasmid was transformed into strain SME 3792, which carries a single copy of an Φ(*virB-lacZ*) promoter fusion, and assayed for transcription activation in comparison with full-length VirF (Table 19). Cells were grown in either Luria Burtani broth (LB) or minimal media (MM) to determine the impact of catabolite repression on expression of *virF* from the *lac* promoter of pHG165. Activation by full-length VirF was lower in cultures grown in LB than in MM, as was expected. The isolated DBD was not capable of activating transcription above the level of the background activation in LB media, and presumably would not activate above basal levels in MM. Therefore, we conclude that residues 2-160 have a function in transcription activation by VirF as the DBD alone is not sufficient for activation.

Screen for Effectors of VirF and Rns. Specific chemical effectors have not been identified for most AraC family virulence regulators. Thus far, the response to environmental stimuli by this group of regulators has been found to occur mostly through indirect regulation, such as temperature-dependent repression by the heat-stable nucleoid

Table 19. *In Vivo* Transcription Activation by VirF and VirF-DBD

	LB	MM
VirF	89.2	968
VirF-DBD	9.9	10.4
Empty Vector	9.6	N/D

β-Galactosidase activity was assayed from a single-copy Φ(*virB-lacZ*) fusion in strain SME3792 and VirF or the VirF-DBD expressed from a pHG165. LB, cells grown in LB+Ampicillin, MM, cells grown in minimal media with glycerol and ampicillin. These values are preliminary data and are the result of a single assay with two replicates.

structuring protein, H-NS (29, 173). However, RegA, an AraC virulence activator in *Citrobacter rodentium,* activates transcription in response to the effector bicarbonate (145). Shortly after this report, the AraC virulence activator ToxT was also reported to activate virulence gene expression in response to bicarbonate (146). Subsequently, ToxT transcription inhibition was shown to be regulated by components of bile, mainly unsaturated fatty acids such as oleic acid (79, 80). The recent full-length crystal structure of ToxT revealed the UFA cis-palmitoleate bound in the effector-binding pocket of the ToxT NTD (79). Another AraC family virulence regulator, UreR, responds to urea to activate transcription (29). Both VirF and Rns share moderate sequence similarity with each of these virulence regulators (Table 20).

Despite low sequence identity of the full-length proteins (Table 20), both the NTD and DBD of ToxT have structural similarity with the respective domains of AraC (79). Based on the structural similarity of the ToxT and AraC NTD's, it is possible that other virulence regulators that respond to effectors also have structural similarity with the AraC NTD. Although we do not have direct structural information for either VirF or Rns, structural predictions from the I-TASSER program (159, 160) support the hypothesis that both proteins may have structural similarity to AraC and ToxT (Figure 17). VirF and Rns (enterotoxigenic *E. coli* (ETEC)) are closely related to one another, share sequence similarity throughout the length of the proteins (Table 20), and can be substituted for one another at their respective promoters (169, 174-176). Given the sequence similarity to virulence regulators known to respond to effectors and structural predictions of VirF and Rns, we hypothesize that VirF and Rns may also respond to a binding of a chemical effector in the effector-binding pocket of the NTD.

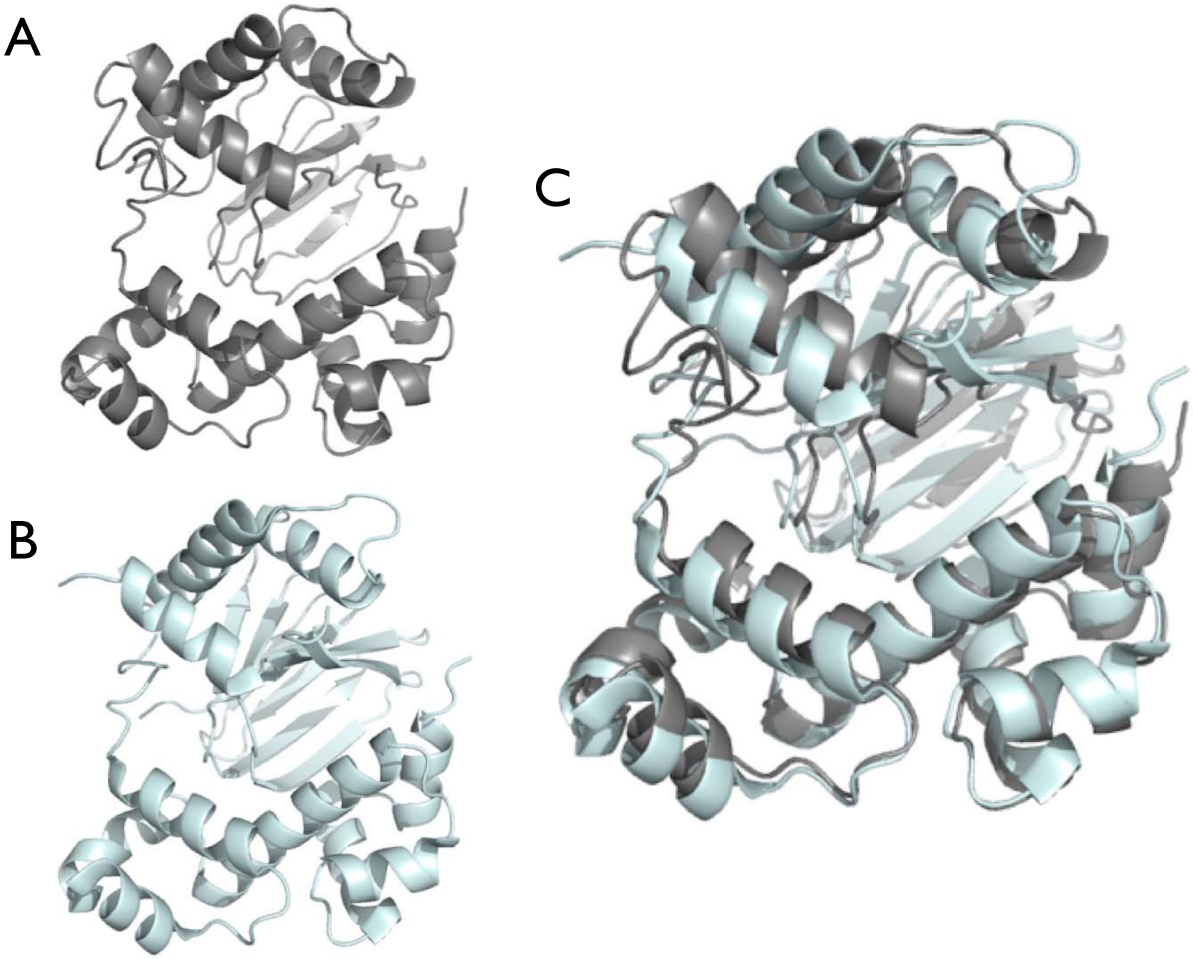


Figure 17. ***In Silico*** **Analysis of the VirF Model Structure.** **A.** Structural prediction for VirF from the I-TASSER server (159, 160). **B.** ToxT crystal structure (79). **C.** Structural alignment of the VirF I-TASSER model (gray) with the ToxT crystal structure (cyan).

	VirF	Rns	AraC	ToxT	RegA
VirF	-	34 (73)	13 (49)	20 (63)	31 (70)
Rns	34 (73)	-	13 (47)	18 (61)	26 (67)
AraC	13 (49)	13 (47)	-	15 (52)	13 (48)
ToxT	20 (63)	18 (61)	15 (52)	-	18 (66)
RegA	31 (70)	26 (67)	13 (48)	18 (66)	-

Table 20. Sequence Identity and Sequence Similarity of VirF and Rns with Other AraC Family Virulence Regulators Regulated by Effectors. Sequence identity and similarity of VirF and Rns to other AraC family virulence regulators that respond to effectors was determined. Values are shown as identity (similarity) in the table above. Identity and similarity scores were generated using the ClustalW multiple alignment program on the Network Protein Sequence Analysis (NPS@) server at: http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html

To test our hypothesis that VirF and Rns respond to chemical effectors, I screened a number of potential physiological effectors that an *Enterobacteriaceae,* such as *Shigella* or ETEC, would likely come into contact with during the course of infection. For *Shigella* to establish infection, it first must pass through the acidic stomach before passing into the small intestine and ultimately into the colon, where it invades colonic and rectal epithelial cells (177, 178). ETEC passes through the stomach to the small intestine to establish infection via adherence to the intestinal epithelial cells (179, 180). During the course of infection, VirF and Rns come into contact with stomach acids, bile salts, and intestinal mucosa, all of which are potential environmental cues to the organism as to its location with the host, and therefore are physiologically relevant sources for effectors of VirF and Rns.

Growth of *Shigella* species in the presence of the bile salt deoxycholate (DOC) or chenodeoxycholate enhanced bacterial invasion of HeLa cells by increasing adherence (181). Furthermore, the factors necessary for enhanced invasion in the presence of DOC are encoded on the *S. flexneri* virulence plasmid (181), which also carries the *virF* gene (129). Based on this report, we hypothesized that perhaps DOC activates expression of VirF, thereby increasing the expression of the type-three secretion system (T3SS) of which VirF is a master regulator (97, 99-102, 106-108). To test this hypothesis, I grew *E. coli* expressing VirF from a plasmid in the presence or absence of various concentrations of DOC within physiological range (1.2mM (5%), 1.5mM, and 2.5mM). I also tested the Rns protein expressed in *E.* coli grown in the presence of DOC to determine if DOC was an effector for Rns (169). The RhaS and RhaR proteins were included in these assays as negative controls as they were not expected to respond to DOC. I found that DOC did not

have a significant effect on VirF or Rns expressed in *E. coli*, or on the negative controls, RhaS and RhaR, at the concentrations tested (Table 21). Although *Shigella* is closely related to *E. coli* (182-184), I wanted to verify that DOC did not affect VirF activity in *S. flexneri* either. Therefore, I grew *S. flexneri* with *virB* fused to the *lacZ* reporter gene [Φ(*virB-lacZ*)] in single copy either in the presence or absence of DOC (Table 22). DOC did not have a significant effect on endogenous VirF activity and therefore we concluded that DOC is likely not an effector of VirF. Enhanced invasion of *Shigella* into host cells in the presence of DOC is not likely due to activation of the VirF protein.

Although DOC does not appear to function as a chemical effector of either VirF or Rns, both *Shigella* and ETEC en counter bile in the course of infection. Furthermore, it had been shown that ToxT activity is inhibited in the presence of bile and bile components (80, 185, 186), specifically UFA's such as oleic acid (79). Therefore, *E. coli* expressing either VirF or Rns from a plasmid was grown in the presence or absence of varying concentrations of bile (0.2%, 0.4%, 0.8%, or 1.0%) or oleic acid (0.02%, 0.04%, 0.08%, or 0.1%). I found that bile did not have an effect on the activity of either activator at the concentrations tested (Table 23). Oleic acid at 0.04% and 0.08% reduced activation by VirF by approximately 10-fold (Figure 18). However, methanol alone reduced activity by approximately 3.5-fold (Figure 18). To determine if the reduced activity was in fact due to oleic acid, a sample of oleic acid at each concentration was re-suspended in one milliliter of methanol and evaporated to 0.1 mL using a speed vacuum centrifuge prior to addition to minimal media. With the methanol evaporated prior to growth, activity levels were not significantly reduced at 0.02% and 0.04% oleic acid and were only reduced by approximately two-fold at concentrations of 0.08% and 0.1% oleic acid (Table 24). It is

Table 21. Effect of DOC on VirF and Rns Transcription Activation in *E. coli*

	VirF	Rns	RhaR (-)rha	RhaR (+)rha
Without DOC	968	4507	310	1126
1.2mM DOC	917	4023	340	1005
1.5mM DOC	1549	4024	397	728
2.5mM DOC	639	5621	310	1141

β-galactosidase activity was assayed from single-copy lacZ fusions with proteins expressed from plasmid pHG165. Assays measured activation by VirF (SME 3793), Rns (SME 3670), or RhaR (SME 3565) of their respective promoters with cells grown in minimal media either with [(+)DOC] or without [(-)DOC] sodium deoxycholate at the indicated concentrations. RhaR was assayed with [(+)rha] or without [(-)rha] L-rhamnose for all conditions. Since DOC had no detected effect on activation, we report results from a single experiment. Values are expressed as Miller Units.

Table 22. Effect of DOC on VirF Transcription Activation in *S. flexneri*

	Without DOC	5% DOC	1.5mM	2.5mM
VirF	4119	5388	5827	5117

β-galactosidase activity was assayed from a single-copy Φ(*virB-lacZ*) in *S. flexneri* grown either with [(+)DOC] or without [(-)DOC] deoxycholate at the indicated concentrations. Cells were grown in LB and values are the result of a two replicates. Values expressed in Miller Units with standard error no greater than 15%.

Table 23. Effect of Bile on VirF and Rns Transcription Activation in *E. coli*

	VirF	Rns
Without bile	968	4507
0.2% bile	676	3848
0.4% bile	1364	4385
0.8% bile	1083	4753
1.0% bile	1291	4823

β-galactosidase activity was assayed from single-copy *lacZ* fusions assays to measure activation by VirF (SME 3793) or Rns (SME 3670) in cells grown with or without bile at the indicated concentrations. Cells were grown in minimal media. Since bile had no detected effect on VirF or Rns activation, we report results from a single experiment. Values are expressed as Miller Units.

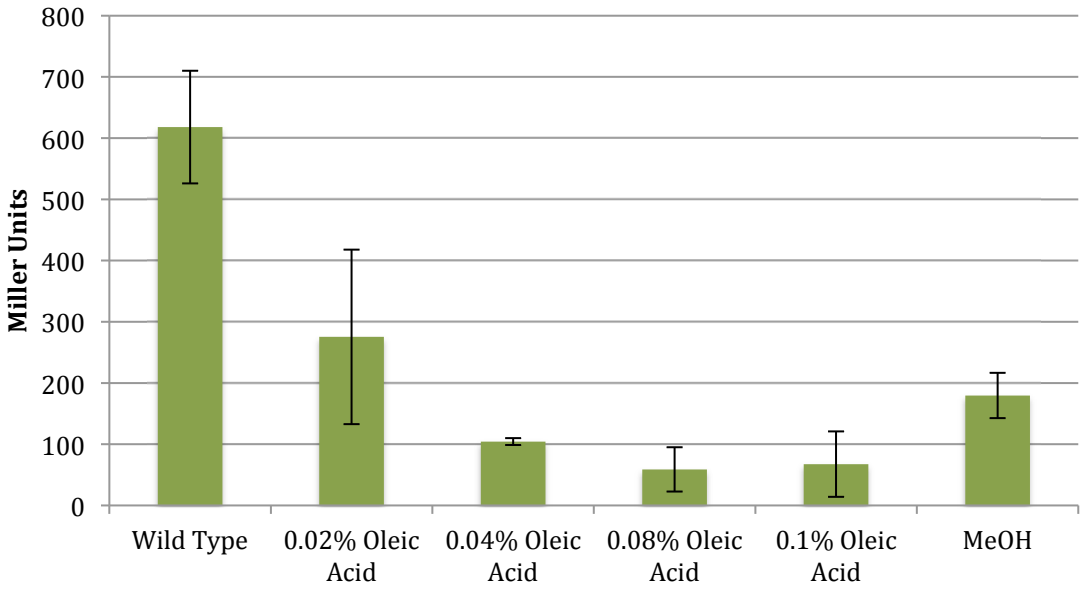


Figure 18. Effect of Oleic Acid on VirF Transcription Activation in *E. coli*.

β-galactosidase activity was assayed from a single-copy Φ(*virB-lacZ*) fusion to measure activation by VirF in cells grown with or without oleic acid at the indicated concentrations. Oleic acid was solubilized in variable volumes of MeOH up to 1mL. Cells were grown in minimal media. Values are the result of two independent replicates and are expressed as Miller Units. Error is the standard error of the mean.

Table 24. Effect of Oleic Acid with MeOH Evaporated on VirF Transcription Activation in *E. coli*

	VirF
WT	683
0.02% Oleic Acid	426
0.04% Oleic Acid	464
0.08% Oleic Acid	353
0.10% Oleic Acid	324
MeOH	N/D

β-galactosidase activity was assayed from a single-copy Φ(*virB-lacZ*) fusion to measure activation by VirF in cells grown with or without oleic acid at the indicated concentrations. Oleic acid was solubilized in MeOH and evaporated to reduce the volume of methanol to which cells were exposed. Cells were grown in minimal media. Evaporated MeOH only control value not determined due to technical error. These values are preliminary data and are the result of a single assay with two replicates. Values expressed as Miller Units.

possible that oleic acid has a slight effect on activation by VirF, however the conclusions regarding the role of oleic acid as an effector of VirF are complicated by the reduced activity of methanol alone. It would be interesting to repeat these results with a solvent other than methanol to determine the direct effect of oleic acid.

The mucosal layer of the intestine is yet another source for a potential effector(s) of VirF. To establish infection, *Shigella* must first penetrate the colonic mucosal layer to gain access to microfold cells (M cells), through which *Shigella* are transcytosed across the epithelial layer (117, 118, 187). The colonic mucosal layer consists predominantly of carbohydrates as well as membrane-bound glycoproteins and high-molecular weight mucin glycoproteins (188-191). In addition to the components of the colonic mucus layer, degradation of plant cell wall polysaccharides and mucins also occurs in the colon (192), further extending the list of potential effectors that VirF may encounter during an infection. A number of mucosal components from the small intestine (fucose, galactose, mannose, galactosamine, and glucosamine) and cecum (gluconic acid, glucuronic acid, and galacturonic acid) have been shown to perpetuate colonization of *E. coli* in the mouse intestine and stimulate *in vitro* growth of *E. coli* (193-196).

To determine the effect of the mucosal carbohydrates glucosamine and glucuronic acid on VirF activation, *E. coli* expressing VirF from a plasmid was grown in the presence or absence of glucosamine or glucuronic acid (2mg/ml, 5mg/ml, 10mg/ml). Neither carbohydrate had a significant effect on the activity of VirF (Table 25). I also performed a screen of common sugars that may be present in colon following digestion of plant cell wall polysaccharides and flax seed oil, which contains multiple fatty acids. For this screen, *E. coli* expressing VirF from a plasmid was spread onto media containing X-gal and

Table 25. Effect of Colonic Mucosal Carbohydrates on VirF Transcription Activation in *E. coli*

	Activation
Wild Type VirF	553
2mg/mL Glucosamine	845
5mg/mL Glucosamine	823
10mg/mL Glucosamine	845
2mg/mL Glucuronic Acid	411
5mg/mL Glucuronic Acid	390
10mg/mL Glucuronic Acid	431

β-galactosidase activity was assayed from a single-copy Φ(*virB-lacZ*) fusion to measure activation by VirF in cells grown with or without glucosamine or glucuronic acid at the indicated concentrations. Cells were grown in minimal media. Since these carbohydrates had no detected effect on VirF activation, we report results from a single experiment. Values are expressed as Miller Units.

Table 26. Effect of Carbohydrate Solutions on VirF Transcription Activation in *E. coli*

Effector Candidate	Inhibition of Rns	Inhibition of VirF
L-Arabinose	+++	+++
Dextran	-	-
D-Fructose	-	-
L-Fucose	+	+
Fumaric Acid	-	-
D-Galactose	+	+
D-Glucose	++	++
D-Lactose	+	+
D-Maltose	+	+
D-Mannose	+	+
D-Melibiose	+	+
L-Rhamnose	+	+
Ribitol	+	+
D-Ribose	+	+
Sorbitol	+	+
Succinic Acid	-	-
Sucrose	-	-
D-Xylose	+	+
Xylitol	-	-
Deoxycholic Acid	-	-
Citric Acid	-	-
Flax Seed Oil	-	-

Strain SME3793 or SME3670 was spread onto media containing X-gal and ampicillin. Sugars were prepared in 20% or 5% solutions and spotted onto the plate containing bacteria. Inhibition was assessed based on the presence of lighter blue or white colonies in comparison with colonies grown without additives. (-) indicates no effect of carbohydrate; (+), (++) and (+++) indicate changes in colony color suggesting decreased activity, (++) and (+++) indicate stronger effect of carbohydrate compared to (+) results. Results are from three independent experiments.

ampicillin. Carbohydrates were prepared in 20% solutions and spotted onto the plate containing bacteria. Inhibition was assessed based on the presence of lighter blue or white colonies in comparison with colonies grown without additives (Table 26). Colonies grown in the presence of arabinose or glucose were lighter blue in color compared to the colonies grown without effector candidates and colonies grown with arabinose showed growth defects. Other sugars also gave lighter blue colonies, but did not have as strong of an impact as either arabinose or glucose. The growth defect seen with arabinose is likely due to deletion of the *araD* gene in the MC4100 parent strain, resulting in accumulation of toxic phosphorylated sugar intermediates and inhibition of growth (197). Furthermore, light blue colonies in the presence of the sugars may have been the result of catabolite repression, and not a result of these sugars acting as effectors of VirF. To test for catabolite repression, strain CSH141 (153), which contains a wild-type lac operon, was streaked in the presence of sugars that decreased indole production by SME3793 or SME3670 on media containing X-gal (198). If a sugar induces catabolite repression, cAMP levels within the cell will decrease, resulting in decreased activation by CRP. A decrease in the amount of active CRP, in turn, results in decreased expression from the *lac* promoter, which drives expression of *virF* and *rns* from plasmid pHG165. Therefore, if indole production is decreased in the presence of a sugar in strain CSH14, it is likely that any reductions in activation by VirF or Rns are due to reduced protein levels and not decreased activation by VirF or Rns. Similar levels of repression were seen in CSH141 as in SME3793 or SME3670 with the exceptions of lactose and sorbitol, which are not expected to induce catabolite repression of the lac operon (199, 200). Flax seed oil had no observed impact on activation by VirF (Table 26).

Table 27. Effect of Sodium Bicarbonate on VirF and Rns Transcription Activation in *E. coli*

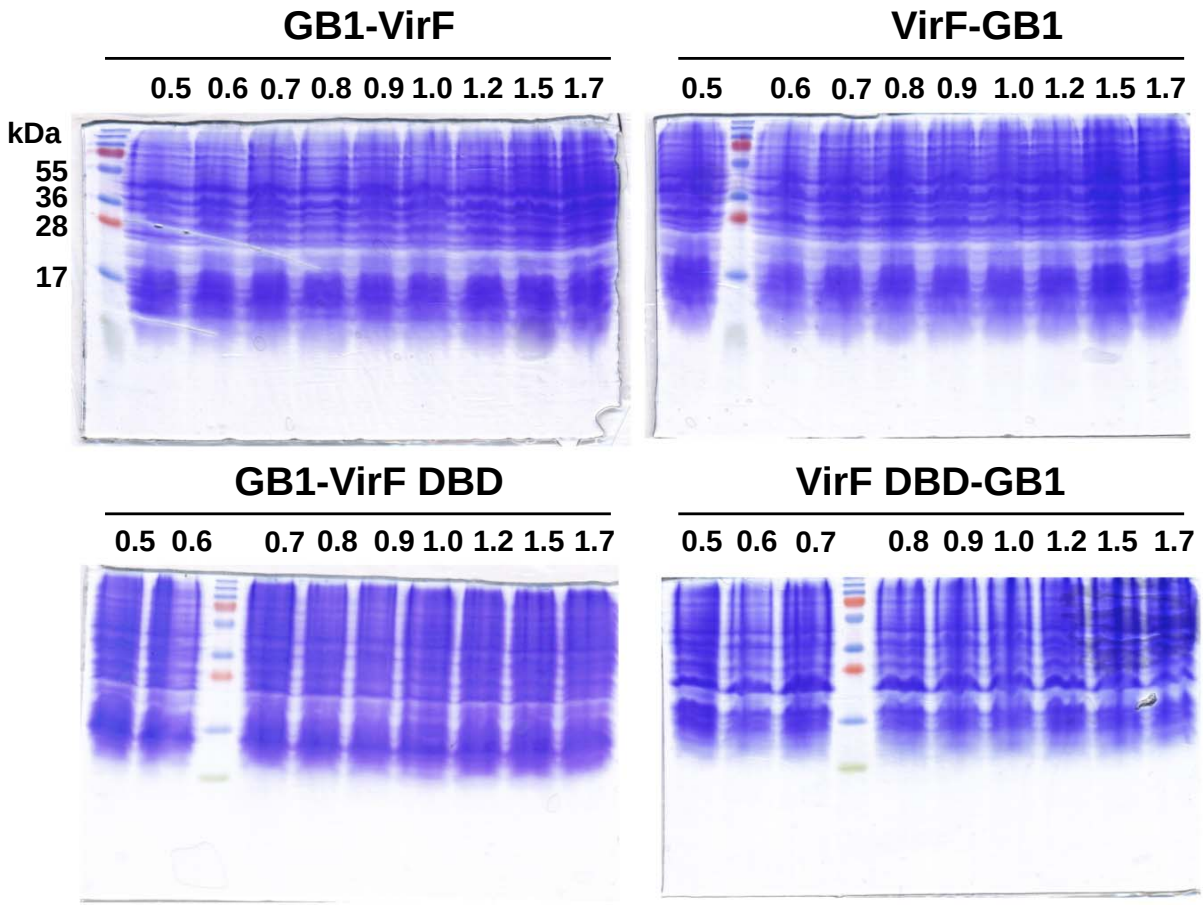
Bicarbonate	% WT Activation			
	VirF	Rns	RhaR (-)rha	RhaR (+)rha
0mM	100±11	100±20	100±2	100±16
10mM	75±8	70±5	102±4	106±15
20mM	50±10	108±22	135±3	160±15
40mM	30±3	65±9	410±1	189±5

β-galactosidase activity was assayed from single-copy lacZ fusions with proteins expressed from plasmid pHG165. Assays measured activation by VirF (SME 3793), Rns (SME 3670), or RhaR (SME 3565) of their respective promoters with cells grown either with [(+)$NaHCO_3$-] or without [(-)$NaHCO_3$-] sodium bicarbonate at the indicated concentrations. Cells were grown in minimal media. RhaR was tested with [(+)rha] or without [(-)rha] L-rhamnose. Values are the result of a two experiments with at least two replicates in each assay. Values expressed in Miller Units with standard error no greater than 22%.

Bicarbonate was tested as a candidate effector for VirF and Rns as bicarbonate had previously been reported to increase transcription activation for RegA and ToxT (145, 146). Additionally, bicarbonate has been shown to affect the expression of virulence-associated genes in many pathogenic bacteria, including, but not limited to, anthrax toxin genes of *Bacillus anthracis* (201), toxic shock syndrome toxin 1 of *Staphylococcus aureus* (202), and genes in the locus for enterocyte effacement (LEE) of EHEC (203). In the intestine, sodium bicarbonate is secreted from the pancreas into the duodenum, neutralizing incoming acidic fluid from the stomach (204) with an approximate concentration in the ileum of 40mM (203, 205, 206).

To determine if bicarbonate does act as an effector of either protein, I grew *E.* coli expressing either VirF or Rns from a plasmid in the presence or absence of various concentrations of sodium bicarbonate, up to physiological concentrations around 40mM (10mM, 20mM, and 40mM). At physiological concentrations (40mM), bicarbonate decreased VirF-dependent activation of a *virB-lacZ* promoter fusion by approximately three-fold (Table 27). Bicarbonate did not have a significant impact on Rns-dependent transcription activation of *cfaA-lacZ*. Although more work is required to fully understand the impact of bicarbonate on VirF-dependent activation, I hypothesize that bicarbonate acts as a repressor of VirF activity.

VirF and VirF^{co} Protein Expression and Toxicity Studies. Biochemical characterization of VirF requires purified, soluble, active protein. The first step towards achieving this goal was to test for overexpression of the protein under inducing conditions. To test for expression levels, SME3750 was transformed with VirF or VirF DBD with either a N- or C-terminal His_6-tagged streptococcal GB1 domain. Samples were induced at various



protein solubility and folding. None of the samples tested showed significant

Figure 19. Overexpression of GB1-Tagged VirF and VirF DBD in SME3750. Expression of N- or C-terminally GB1 tagged VirF or VirF DBD was induced at various OD_{600} readings (indicated above each lane) with 1.0mM IPTG at 15°C overnight. Cells were heat lysed and electrophoresed on a 12% SDS-polyacrylamide gel and stained with Coomassie blue. The expected size of each fusion protein is as follows: N- or C-terminally tagged VirF, 38 kDa; N- or C-terminally tagged VirF DBD, 20 kDa. Because over expression of these fusion proteins was not significant, only a single experiment was performed.

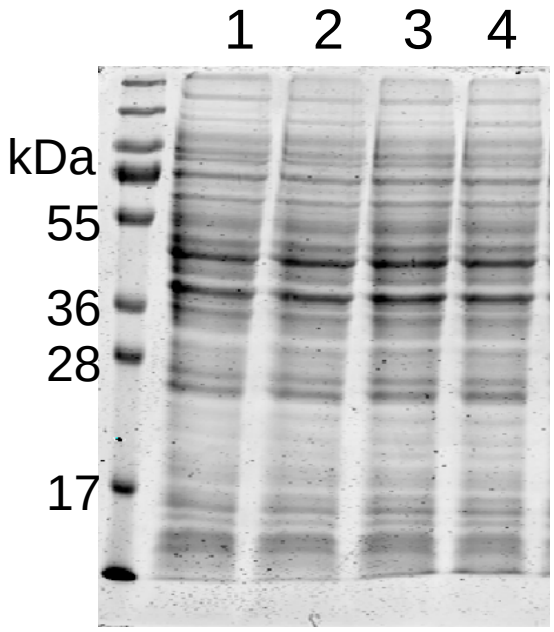


Figure 20. Overexpression of GB1-Tagged VirF and VirF DBD in pLysS (DE3).

Expression of N- or C-terminally GB1 tagged VirF or VirF DBD was induced at an OD_{600}=1.0 with 0.1mM IPTG at 15°C overnight. Cells were lysed by three freeze-thaw cycles and electrophoresed on a 12% SDS-polyacrylamide gel and stained with Coomassie blue. The expected size of each fusion protein is as follows: N- or C-terminally tagged VirF, 38 kDa; N- or C-terminally tagged VirF DBD, 20 kDa. Lane 1: GB1-VirF; Lane 2: VirF-GB1; Lane 3: GB1-VirF DBD; Lane 4: VirF DBD-GB1. Because over expression of these fusion proteins was not significant, only a single experiment was performed.

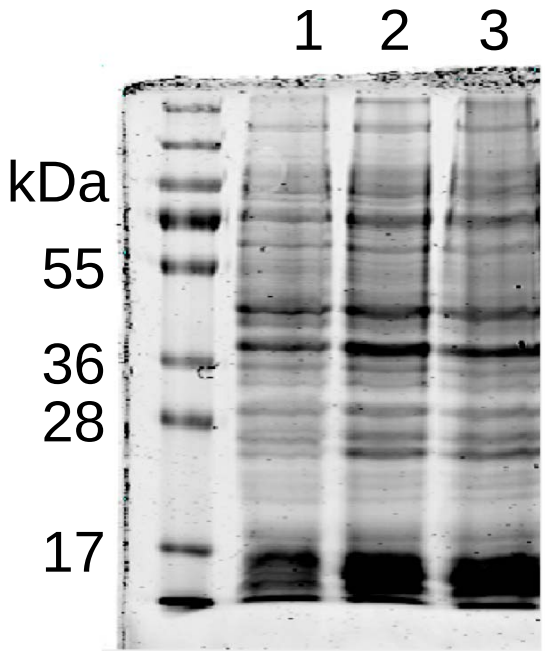


Figure 21. Overexpression of GB1-Tagged VirF and VirF DBD in Rosetta2 (DE3). Expression of N- or C-terminally GB1 tagged VirF or VirF DBD was induced at an OD_{600}=0.8 with 1.0mM IPTG at 15°C overnight. Cells were lysed by sonication and electrophoresed on a 12% SDS-polyacrylamide gel and stained with Coomassie blue. The expected size of each fusion protein is as follows: N- or C-terminally tagged VirF, 38 kDa; N- or C-terminally tagged VirF DBD, 20 kDa. Lane 1: GB1-VirF; Lane 2: VirF-GB1; Lane 3: GB1-VirF DBD. Transformants of Rosetta2 (DE3) carrying pDZ3/*virF* DBD were not obtained, thus overexpression of VirF DBD-GB1 was not tested in this strain. Because over expression of these fusion proteins was not significant, only a single experiment was performed.

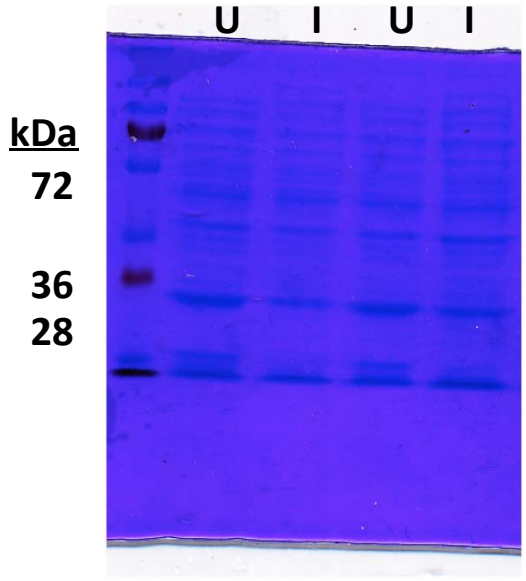


Figure 22. Overexpression of VirF- His_6 in pLysS (DE3). Expression of VirF with a C-terminal histadine tag was induced at an OD_{600}=1.0 with 0.1mM IPTG at 37°C for 3 hours. Cells were heat lysed by boiling for 10 minutes and electrophoresed on a 12% SDS-polyacrylamide gel and stained with Coomassie blue. The expected size of the fusion protein is 30 kDa. U: uninduced control; I: induced sample. Because over expression of the VirF-His_6 fusion protein was not significant, only a single experiment was performed.

OD_{600} readings (0.5-1.7) to determine conditions that resulted in any significant expression. Cells were induced overnight with 1mM IPTG at 15°C with shaking to increase overexpression of either full-length protein or of the VirF-DBD (Figure 19). Expression was then tested in BL21 (DE3) pLysS and Rosetta2 (DE3). The BL21 (DE3) pLysS cell line carries the pLysS plasmid, which carries the gene encoding the T7 lysozyme. T7 lysozyme suppresses basal expression from the T7 promoter driving expression of the VirF fusion proteins under non-inducing conditions. The Rosetta2 (DE3) cell line supplies tRNA's for seven rare codons in *E. coli* and also carries the pLysS plasmid. Overexpression was not detected in either of these strains (Figures 20 and 21). Another group had reported successful purification of a VirF fusion protein (VirF-His_6) (108). Therefore I constructed a VirF-His_6 fusion protein and again tested for expression using the previously described protocol (108), however expression was very poor, as seen with the GB1-tagged constructs (Figure 22).

Due to the poor expression of GB1-tagged VirF and the VirF-DBD and of VirF-His6, I considered the possibility that overexpression of VirF may be toxic to the cell, leading to plasmid instability. I therefore performed plasmid stability testing per the protocol included in the pET System Manual (Novagen Billerica, MA) since the pDZ1 and pDZ3 plasmids are derivatives of pET21a. Results did not indicate plasmid instability as an explanation for low expression of the VirF constructs. To test for plasmid instability, the expression strain carrying the target plasmid was spread on four plates with the indicated additives: Kanamycin (allows growth of all viable cells of strain SME3750), Kanamycin and Ampicillin (only cells that have maintained the plasmid can grow), Kanamycin and IPTG (only cells that have lost the plasmid or the ability to express the target protein can grow),

Kanamycin with Ampicillin and IPTG (only cells that have maintained the plasmid, but cannot express the target protein can grow).

Equal growth was seen on plates containing kanamycin (selection for SME3750) and both kanamycin and ampicillin (selection for the pDZ1/pDZ3 plasmids), indicating that most of the cells maintained the plasmid. Very little growth was seen when cells were plated with kanamycin and IPTG, which is expected since nearly all of the cell's resources are converted to target gene expression, resulting in very little cell growth after induction. This suggested that very few cells had lost the plasmid and continued to grow after induction. Lastly, no growth was observed on the plate containing kanamycin, ampicillin, and IPTG. Again, this is expected as only cells that have maintained the plasmid are resistant to ampicillin, but little growth occurs due to expression of the T7 RNA polymerase and conversion of cellular resources for target gene expression.

Since the plasmid stability test indicated that expression of VirF is non-toxic to host cells, I further examined the DNA sequence of the gene encoding VirF. The *virF* gene encodes a large number of rare codons in *E. coli* in addition to low G/C content, which could explain low levels of expression. I therefore ordered a codon-optimized version of the *virF* gene, which changes rare codons to those more frequently utilized in *E. coli* without altering the amino acid sequence of the protein. The codon-optimized gene, *virF*^{co}, was cloned into the pMAL-C2x expression vector for expression. I chose a maltose binding protein (MBP) tag since an MBP-VirF fusion protein has been successfully expressed and purified (106). I was able to achieve sufficient expression of this construct in strain KS1000 using the methods suggested in the pMAL protein fusion and purification manual (New England Biolabs Beverly, MA).

Purification of Active Soluble MBP-VirF^{co}. MBP-VirF^{co} was purified as previously described (157) (Figure 23). Electrophoretic mobility shift assays (EMSA's) were performed to test the DNA-binding activity of MBP-VirF^{co}. The protein was able to bind to DNA containing either the predicted binding sites at the *virB* promoter (the native binding site for VirF) or the rns binding site at the *cfaA* promoter (Figure 24). The *cfaA* promoter was tested as it is recognized by Rns (207) and Rns and VirF have been shown to be interchangeable (176), therefore I hypothesized that it was likely that MBP-VirF^{co} would bind to this fragment of DNA. Since MBP-VirF^{co} shifted DNA, and the amount of shifted DNA decreased with protein concentration, I concluded that MBP-VirF^{co} is well folded and active and therefore is sufficient for further characterization.

Dimerization Studies of MBP-VirF^{co}. Various studies have indicated that there are likely multiple VirF binding sites upstream of VirF regulated promoters, suggesting oligomerization of VirF (106, 108, 172, 208). Furthermore, protein-protein interaction is supported by the isolation of two dominant-negative variants in a mutagenesis study by Porter and Dorman (172). Due to the lack of published biochemical studies characterizing the oligomerization state of VirF, I performed some preliminary studies on the dimerization of VirF using the MBP-VirF^{co} fusion protein. Given the evidence that dimerization may be required for activation by VirF (172), it is unlikely that MBP blocks dimer formation as the fusion protein was able to activate transcription at *virB* in the same way the native *virF* gene product does (106).

DNA footprinting and deletion analysis of the *virB* promoter indicated that the region from -110 (relative to the transcription start site) to -59 is required for maximal activation by VirF (106, 208). While this region has been identified as requisite for VirF-

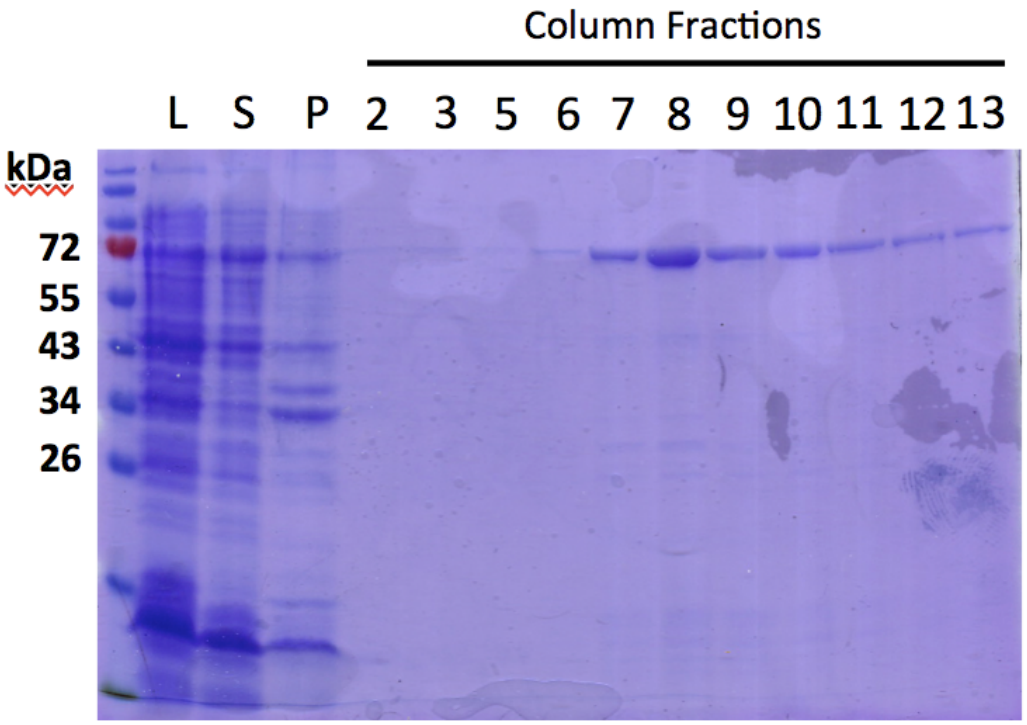


Figure 23. Expression and Purification of MBP-VirF^{co}. Column fractions from purification of MBP-VirF^{co} electrophoresed on a 12% SDS-polyacrylamide gel and stained with Coomassie Blue. Fractions were eluted with 20mM Tris-HCl, pH 7.4, 0.2M NaCl, 1mM EDTA, 1mM DTT with 10mM Maltose. L, sonicated lysate; S, supernatant fraction of sonicated lysate; P, pellet fraction of sonicated lysate. Column fractions are numbered according to the order in which they were eluted. MBP-VirF^{co} was expressed from pMAL-Cx in strain KS1000.

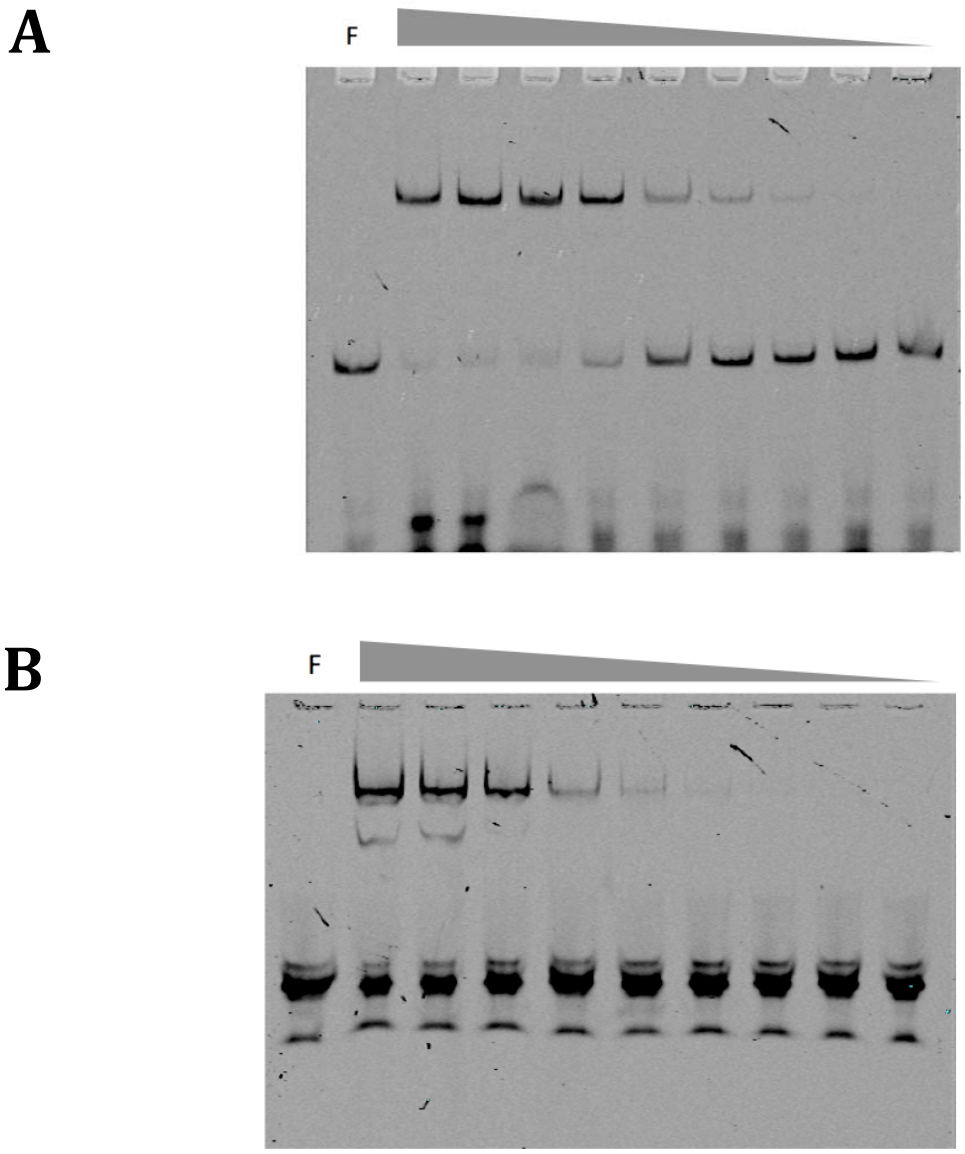
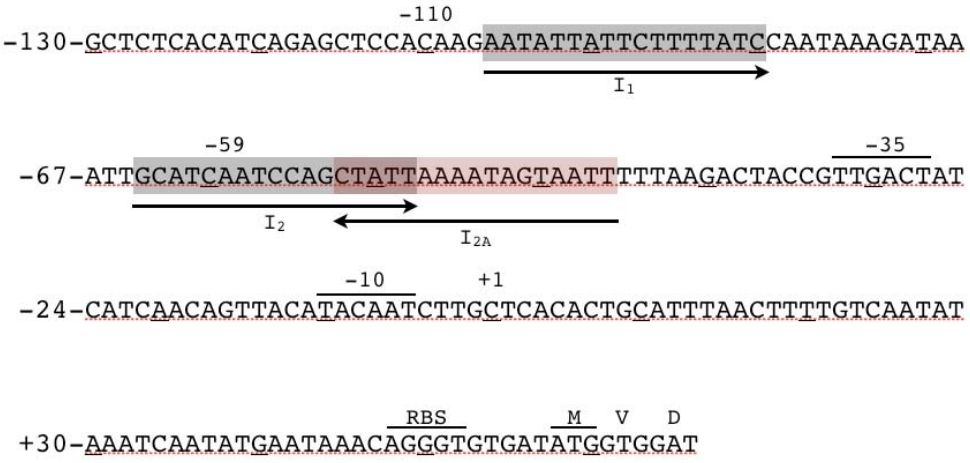


Figure 24. Electrophoretic Mobility Shift Assays to Test DNA Binding Activity of MBP-VirF^{co}. A. The DNA fragment containing the *virB* promoter was generated using the LEUGO procedure with oligos 3371, 3372, and 3373 with oligo 3371 carrying an AminoC6+DY682 infrared tag. The highest MBP-VirF^{co} concentration was 0.64mg/ml and dilution steps were two-fold. F, free DNA. **B.** The DNA fragment containing the *cfaA* promoter was generated using oligos 3160 and 3159, both AminoC6+DY682 labeled. The highest MBP-VirF^{co} concentration was 1.3mg/ml and dilution steps were two-fold. F, free DNA. Images representative of two independent experiments.

dependent transcription of *virB*, the specific sites within this region to which VirF have not been farther refined. To more precisely locate the VirF binding sites, I examined the upstream regions of *virB* using a binding logo generated from 29 known Rns binding-sites (George Munson, personal communication). The Rns binding logo provided a logical starting point for identification of VirF binding sites as VirF can activate transcription of Rns-dependent promoters (176). Using the Rns logo, I identified two likely VirF binding sites (I_1 and I_2) in the region between -110 and the -35 element at the *virB* promoter (Figure 25B). The putative sites are arranged as a direct repeat, are each 17 bases long and separated by 15 bases. The size and spacing of the putative VirF binding sites are consistent with the RhaS, RhaR and AraC binding sites (92, 94, 209). An alternative I_2 site (I_{2A}) was also identified, but is separated from the putative I_1 by 27 bases (Figure 25A). A separation of 27 bases is greater than the maximum separation at which AraC was capable of binding to both half-sites (210). I therefore hypothesize that it is unlikely that I_{2A} is a true binding site for VirF. To determine if VirF binds to the putative I_1 and I_2 sites (*virB*-I_1I_2), I designed oligonucleotides containing both I_1 and I_2, or only the upstream putative I_1 site (*virB*-I_1). I hypothesized that if VirF forms a dimer in solution, that both putative half sites would be required for binding to DNA. I found that MBP-VIrF^{co} does bind to DNA containing both putative half sites, but does not bind to DNA containing only I_1 (Figure 26). While VirF does not bind to the oligo containing only I_1 at *virB*, VirF is able to bind to the *cfaA* promoter, which also only contains a single binding site. Since the consensus binding sequence for VirF is as of yet unknown, it is unclear if the sequence at the *cfaA* promoter is stronger than the sequence at *virB,* which could allow a VirF monomer to bind more tightly, or with higher affinity, than to *virB*-I_1, resulting in a shift in the DNA. Future experiments

A



B

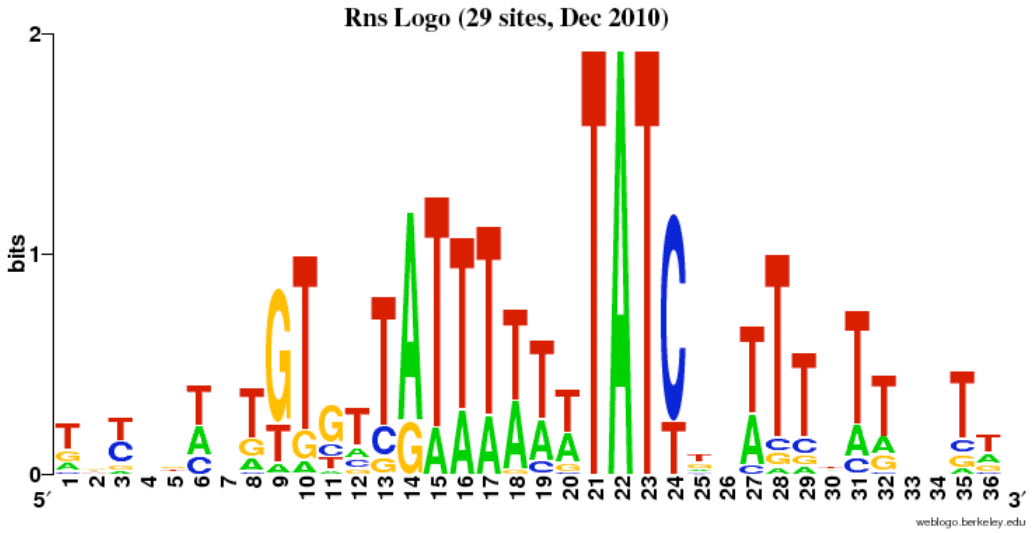


Figure 25. Predicted VirF Binding Sites at the *virB* Promoter. A. DNA sequence of the upstream promoter region of *virB*. Nucleotides are numbered from the transcription start site (+1). Deduced -10 and -35 sequences, putative ribosome binding site (RBS), and transcription start codon are indicated by lines above the sequence. Putative VirF binding sites are shaded in gray and labeled. An alternative I_2 site (I_{2A}) is shaded in pink and labeled. Arrows below the sequence indicate the orientation of the putative repeats. **B.** Rns binding-site logo (Personal Communication, George Munson). Consensus of 29 known Rns binding sites. Height of each nucleotide is proportional to its frequency and the information content of that position. Positions 8-24 used to identify putative VirF binding sites as this region has the highest similarity to known AraC family binding sites (92, 94, 209).

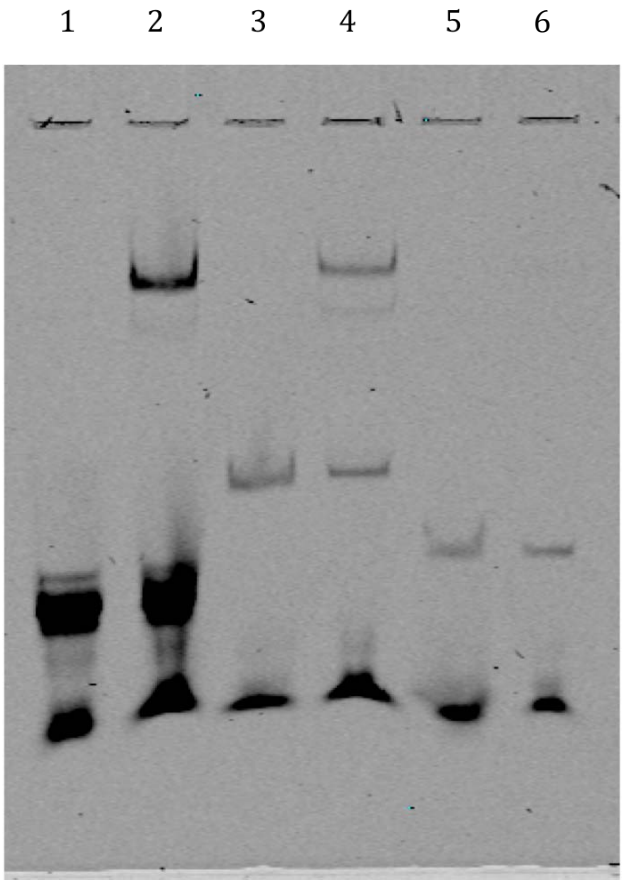


Figure 26. VirF Binding Sites at *cfaA* and *virB*. Final concentration of MBP-VirF^{co} was 2.98μM. Representative EMSA gel image. Lanes 1, 3, and 5 Free DNA; Lanes 2, 4, and 6 with MBP-VirF^{co}. Lanes 1 and 2, the DNA fragment containing the *cfaA* promoter was generated using oligos 3160 and 3159, both AminoC6+DY682 labeled; Lanes 3 and 4, the DNA fragment containing both the I_1 and I_2 half sites at the *virB* promoter was generated using the LEUGO procedure with oligos 3371, 3372, and 3373 with oligo 3371 carrying an AminoC6+DY682 infrared tag; Lanes 5 and 6, the DNA fragment containing the I_1 half-site at the *virB* promoter was generated using the LEUGO procedure with oligos 3371, 3374, and 3375 with oligo 3371 carrying an AminoC6+DY682 infrared tag. Two independent assays were performed for this experiment.

should test binding to DNA containing only the I_2 site to determine if VirF can bind as a monomer to this site, but not to I_1.

To investigate the oligomeric state of MBP-VirF^{co} in solution, I used size exclusion chromatography to separate monomeric protein from larger oligomers. Freshly purified protein was applied to an S200 10/300 analytical gel filtration column pre-equilibrated with Elution Buffer (see Materials and Methods for recipe). A Bio-Rad Gel Filtration standard was also applied to this column, generating a standard curve to determine the size of any peaks of MBP-VirF^{co}. The first peak eluted with an approximate molecular weight of 1,941 kDA, which likely corresponds to aggregates of the fusion protein. The second peak eluted with a size of approximately 184kDa and contained 26% of the total protein (Figure 27). This peak is consistent with the expected size of an MBP-VirF^{co} dimer (145.4kDa) and was symmetrical, indicating a single species was present. Six percent of the protein eluted in a peak of approximately 41.5kDa, which is consistent with the expected size of MBP (42.7 kDa). Several small peaks eluted after the putative MBP peak, but these peaks had molecular weights of approximately 5.7 kDa or less. A peak consistent with the size of an MBP-VirF^{co} monomer (72.7 kDa) was not observed. These preliminary results support that this protein is a dimer in solution, rather than dimerizing upon binding to DNA as has been previously suggested (172).

Electrophoretic Mobility Shift Assays with MBP-VirF^{co} and SE-1. The small-molecule inhibitor SE-1 inhibits DNA binding RhaS and RhaR (147, 148). Results indicated that SE-1 likely binds to the conserved DBD, thus blocking the ability to bind DNA and suggesting that SE-1 may inhibit activation by additional AraC family activators (147, 148). *In vivo*, SE-1 inhibited activation of a *virB-lacZ* transcriptional fusion by VirF (147, 157). To

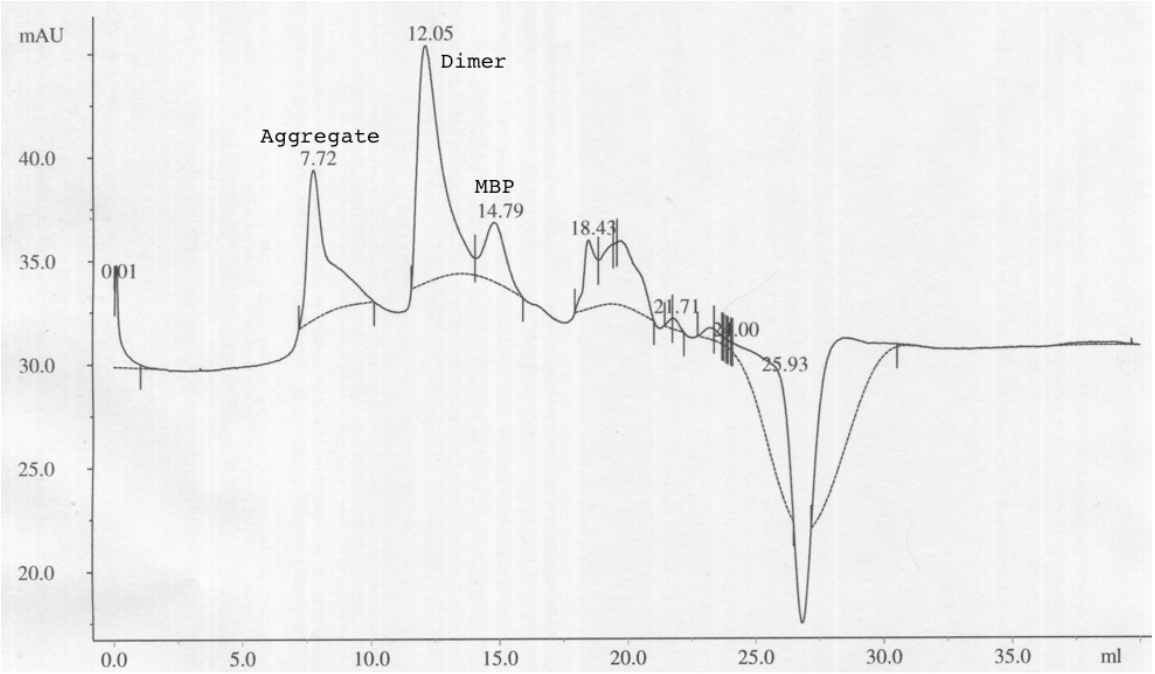


Figure 27. Gel Filtration of MBP-VirF^{co} to Determine the Oligomeric State in Solution. Chromotography trace of MBP-VirF^{co} after gel filtration using a Superdex200 10/300 column for analysis. The column was pre-equilibrated in Elution Buffer. A Bio-Rad gel filtration standard re-suspended in Elution Buffer was used to generate a standard curve to calculate molecular weight of peaks. The X-axis represents the elution volume while the Y-axis represents the UV absorbance. The elution volume is labeled for each peak. Preliminary results reported from one independent experiment.

determine if the compound blocks the ability of VirF to bind DNA *in vitro*, as with purified RhaS and RhaR (147, 148), I performed Electrophoretic Mobility Shift Assays (EMSA) with purified MBP-VirF^{co} and DNA containing the VirF binding sites at the *virB* promoter in the presence or absence of SE-1. Preliminary results indicated that SE-1 at a concentration of 1.3mM inhibited DNA binding by a range of 50 to 100 percent, while inhibition at 0.65μM SE-1 ranged from 50 percent to no inhibition (Figure 28). Final assays performed by Veerendra Koppolu showed that SE-1 inhibited VirF completely at 1.3μM and 0.65μM with approximately 50% inhibition at 0.32μM (147). In my EMSAs, the final concentrations of DNA and protein were at 5nM and 700nM while the final concentrations of DNA and protein were 2nM and 300nM in the published assays (147). The elevated protein concentration in my assays may have shifted the equilibrium of protein with and without inhibitor bound such that inhibition by SE-1 was artificially reduced. Reduction of the protein concentration resulted reproducible inhibition at lower concentrations. We concluded that SE-1 is an inhibitor of VirF, and has potential for development as an AraC family-specific antimicrobial drug (147).

Repression of VirF at *nlpA*. As VirF has been shown to be interchangeable with Rns (176), I tested the ability of VirF to repress transcription at the *nlpA* promoter, which is normally repressed by Rns (211). If VirF can repress transcription of *nlpA*, then transcriptional fusions of the *nlpA* promoter with *lacZ* could be used to separate VirF variants with DNA binding defects from variants with defects in processes required for transcription activation, which could provide more insight into how VirF activates transcription. To test the ability of VirF to repress transcription at *nlpA*, I performed β-galactosidase assays in strain SME3613, which carries a transcriptional fusion of *nlpA*

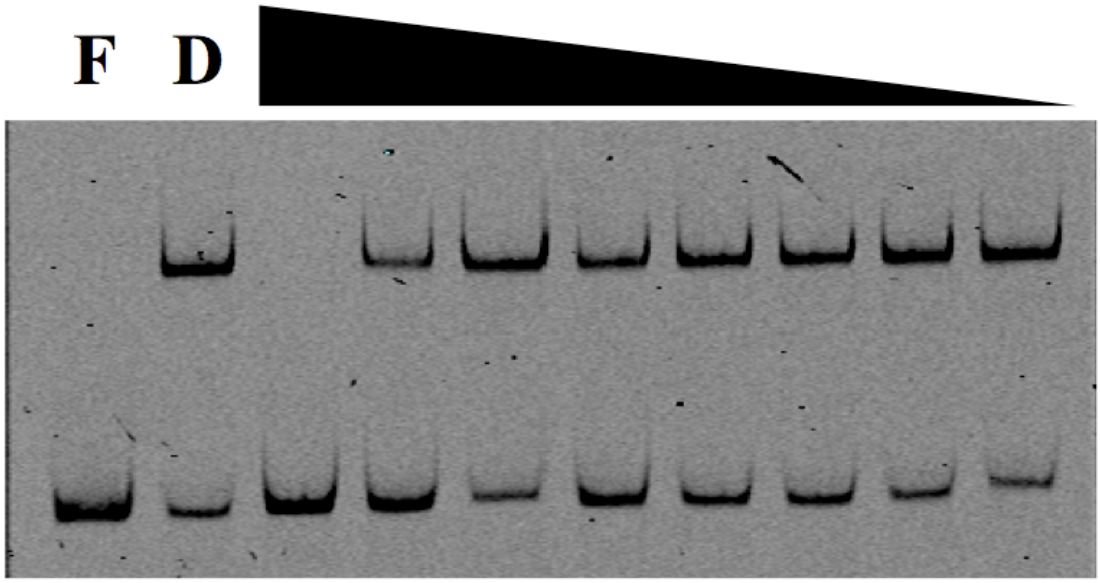


Figure 28. Electrophoretic Mobility Shift Assay to Measure Inhibition of *In Vitro* DNA Binding by MBP-VirF^{co} in the Presence of SE-1. Representative image of EMSA gels. The DNA fragment containing the *virB* promoter was generated using the LEUGO procedure with oligos 3371, 3372, and 3373 with oligo 3371 carrying an AminoC6+DY682 infrared tag. MBP-VirF^{co} was used at 700nM with DNA at 5nM. F, free DNA. D, DMSO only control. Inhibitor concentrations ranged from 1,300 to 10μM with two-fold serial dilutions. Results are from ten replicates.

with *lacZ* Φ(*nlpA-lacZ*) in single copy, transformed with plasmid pHG165/*virF* or pHG165/*rns*. Preliminary results indicated that VirF likely was able to repress transcription of *nlpA*, although not to the same extent as Rns (Figure 29). Despite lesser inhibition of *nlpA* by VirF than by Rns, the *nlpA-lacZ* fusion will be useful in future work to further define the mechanism of transcription activation by VirF.

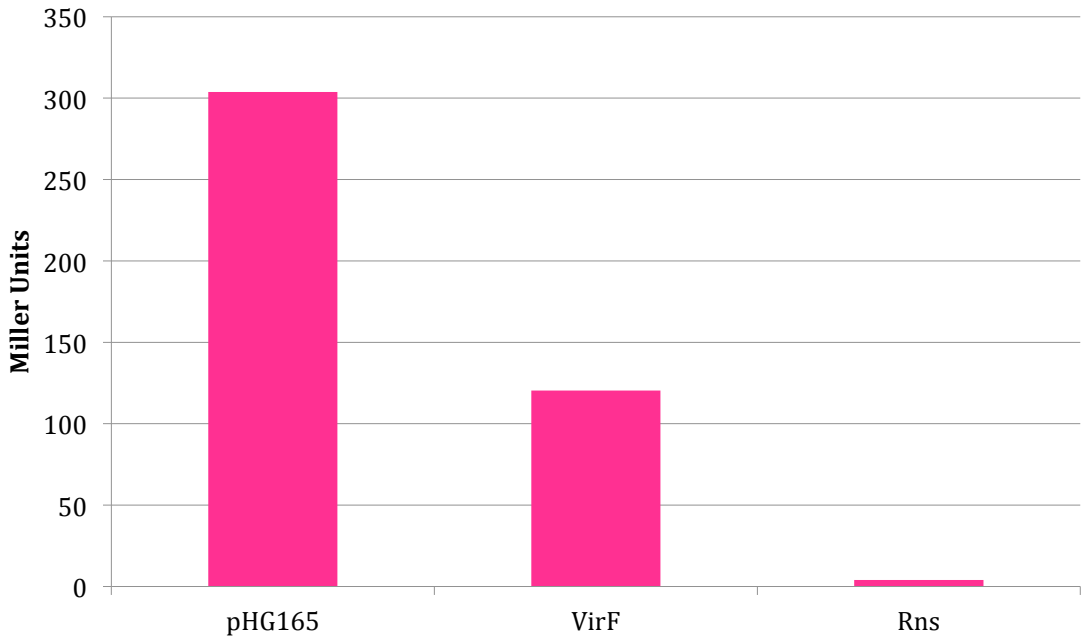


Figure 29. Repression of *nlpA-lacZ* by VirF. β-galactosidase activity was assayed from a single-copy Φ(*nlpA-lacZ*) fusion to measure repression by VirF or Rns. Cells were grown in minimal media. Preliminary results from one replicate. Values are expressed as Miller Units.

Chapter V: Discussion

Model for the RhaR allosteric effector response. Drawing from our mutagenesis studies and previous work on RhaR in addition to parallels with other AraC-family proteins, we propose the following model for allosteric L-rhamnose signaling in RhaR, first focusing on the (-)rhamnose state. We propose that, in the absence of L-rhamnose, the RhaR Arm region (NTD, residues 35-42) contacts the AS2 region (DBD, residues 261-270). These interdomain contacts are proposed to be at least one component of the mechanism that influences the conformation and/or dynamics of RhaR to reduce DNA binding and contacts with RNA polymerase (likely σ^{70} (37)) relative to the (+)rhamnose state. As a result, RhaR activates transcription to only low basal levels (-)rhamnose. We previously found that substitutions at RhaR T279 resulted in increased activity (-)rhamnose, and proposed (based on the predicted position of T279 between RhaS and RhaR residues that contact σ^{70} (37, 38)) that, (-)rhamnose, T279 may interfere with effective RhaR contacts with σ^{70} (144).

Our current results support the hypothesis that the inhibitory contacts involving (at least) RhaR L35 and H269 may be involved in positioning T279 to interfere with RhaR-σ^{70} contacts (-)rhamnose. Structural models of RhaR predict that residues in the Arm and AS2 are in position to contact one another and are fairly near T279 (Figure 30), providing a possible mechanism by which the empty ((-)rhamnose) effector-binding pocket could influence the conformation of T279, and thereby reduce RNA polymerase contacts. In support of this model, the ToxT residue that aligns with RhaR H269 (ToxT T231), is adjacent to residue K230, which has been predicted to be integral to the formation of the "closed" non-activating conformation in ToxT (79).

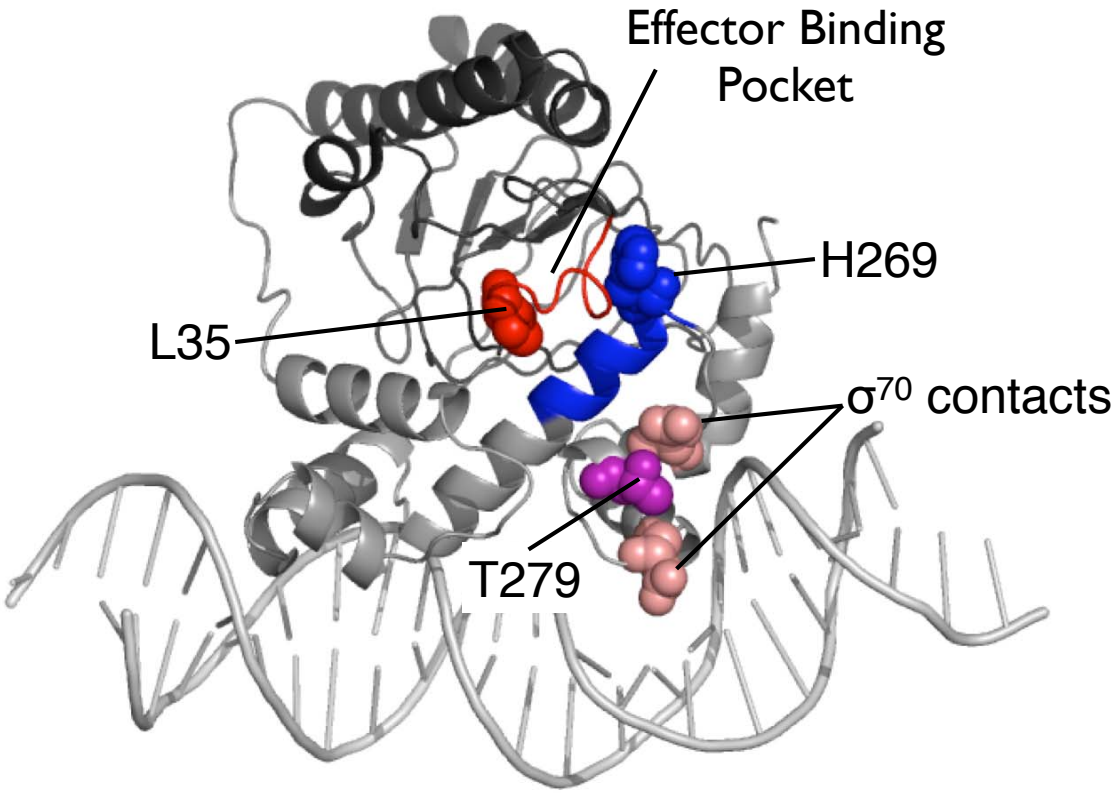


Figure 30. ***In silico*** **Analysis of the RhaR Structure.** Predicted structure of RhaR without the Extension and (-)rhamnose modeled onto DNA from the MarA co-crystal (43). NTD shown in dark gray, DBD in light gray. Model predicts residues L35 in the Arm (red spheres) and H269 in AS2 (blue spheres) are in a position to make contact and are positioned near T279 (purple spheres). We propose a model where the empty ((-)rhamnose) effector binding pocket could influence the position of T279 to interfere with RNAP-σ^{70} contacts (pink spheres), inhibiting transcription activation.

In the presence of L-rhamnose, all of the variants at L35 (Arm) and H269 (AS2, with the exception of a proline substitution in an α-helix) had wild-type activity. This indicated that the contacts between Arm and AS2 are lost (+)rhamnose, resulting in a relief of the proposed (-)rhamnose inhibition. We propose that loss of the Arm – AS2 contacts results in a conformational change in the RhaR DBD that alters the conformation of T279 and allows RhaR (including residue D276 (37)) to effectively contact RNA polymerase and activate transcription.

The RhaR-DBD alone was previously found to be capable of binding to DNA but barely able to activate transcription (90, 93, 94). This suggests that, in addition to the proposed (-)rhamnose inhibitory contacts identified in this work, the RhaR NTD makes stimulatory interdomain contacts (+)rhamnose for RhaR to effectively activate transcription. However, among the 157 variants isolated at 31 positions, we did not isolate a single variant in the Arm, AS1, AS2, or C-HTH regions that decreased RhaR activation (+)rhamnose while maintaining wild-type activity (-)rhamnose (the phenotype expected for loss-of-residue-function variants at positions that make stimulatory contacts (+)rhamnose). While it is possible that RhaR does not make stimulatory contacts (+)rhamnose, the phenotype of the RhaR-DBD alone argues against this hypothesis. Thus, it seems more likely that interdomain stimulatory contacts are made in regions of RhaR other than the Arm, AS2 or C-HTH regions. Alternatively (or in addition), there may be redundancy among the stimulatory contacts (+)rhamnose such that single substitutions do not exhibit loss of (+)rhamnose activity. Significant redundancy among the RhaR stimulatory contacts (+)rhamnose but not the inhibitory contacts (-)rhamnose might suggest that there is a larger number of interdomain contacts (+)rhamnose than

(-)rhamnose, consistent with the requirement of both domains for efficient transcription activation.

RhaR AS2 Region Inhibits Activity in the Absence of L-rhamnose. Site-directed random mutagenesis in the DBD region predicted to be in contact with the RhaR Arm was performed to determine the role of these residues in allosteric signaling in RhaR. These residues are predicted to be located on the DNA distal surface of the C-terminal subdomain of the DBD (43) in a region we call Allosteric Site 2 (AS2). RhaR AS2 residues align with AraC and ToxT residues predicted to be in contact with their respective NTDs to maintain their "closed", non-activating states (68, 79). Many of the variants in RhaR AS2 had phenotypes that suggest protein instability (Table 9, Down and Down (-)). However, with the exception of residue L267, variants in MarA and SoxS (both single domain proteins, DBD only) at positions aligning with RhaR AS2 did not have stability defects and had little or no effect on activity (163, 212). This supports the hypothesis that residues in RhaR AS2 may make interdomain contacts (consistent with the ToxT structure (79)), and that these contacts stabilize the RhaR protein structure.

I identified RhaR AS2 residue H269 as having a potential role in L-rhamnose signaling. Two variants at this position (H269S and H269K) had increased activity (but not elevated protein levels) (-)rhamnose and wild-type activity (+)rhamnose (Table 9). Although only two of the six variants at position 269 had increased activity (-)rhamnose, the other four variants had phenotypes that may further indicate a role for this residue in interdomain contacts (discussed below). Increased activity (-)rhamnose is consistent with two variants at the aligned position: D247V in XylS, exhibited increased basal activity in its inactive state (-)effector, but did not increase protein levels (83), and variant T241E in

AraC, which had increased basal activity without strengthening the interaction of AraC with RNA polymerase (68). We therefore propose that H269S and H269K may be loss-of-residue-function variants and that RhaR H269 may inhibit RhaR activity (-)rhamnose, again similar to the AraC "light-switch" mechanism (64-68).

RhaR Arm-AS2 Contacts Result in Inhibition in the Absence of L-Rhamnose. Identification of residues in the RhaR Arm (L35) and AS2 regions (H269) that appear to inhibit RhaR activity (-)rhamnose suggested that RhaR may use a mechanism for allosteric signaling that is similar to that of AraC (although not involving DNA looping) (64-68). In AraC, (-)arabinose the Arm is predicted to contact the DBD at the region we refer to as AS2 (68). This contact between the Arm and AS2 inhibits AraC from binding to the DNA sites from which it activates transcription (68). Additionally, structural predictions of the RhaR protein indicated that residues in the Arm are positioned near, and likely contact, residues in AS2. Thus, we hypothesized that, (-)rhamnose, RhaR L35 might contact H269 to maintain the low basal activity of RhaR in the absence of effector.

To test this hypothesis, we examined the potential interaction of L35 with H269 (-)rhamnose by screening for intragenic second-site suppressor mutations that restored (-)rhamnose activation to wild-type, non-activated levels. Although second-site suppressors are most commonly associated with restoring activity to deleterious mutations, the same concept has previously been used to identify second-site suppressors of a mutation with increased function (213). We identified H269V as a second-site suppressor of L35K, supporting the hypothesis that L35 and H269 may interact (-)rhamnose. As is common for second site suppressors, the two positions swapped physicochemical properties, going from hydrophobic to positively charged in one case

(L35K) and from positively charged to hydrophobic in the other case (H269V).

We hypothesize that the L35K substitution results in loss of inhibitory contacts (-)rhamnose, and that these inhibitory contacts are restored by the addition of the H269V substitution. Thus, the activity of RhaR (-)rhamnose is returned to its lower, non-activating level (Table 14). The findings that H269V had wild-type activity (-)rhamnose as a single substitution, and that second site suppressors could not be identified for L37R (as a predicted gain-of-function variant, L37 is not predicted to make inhibitory contacts with the DBD in wild-type RhaR), both support the hypothesis that L35 and H269 make contacts (-)rhamnose that contribute to maintaining the low basal level activity of RhaR.

However, the predicted contact propensity between leucine and histidine in the wild type protein (or lysine and valine in the second site suppressor variant) is lower than expected for a true interaction (167). Therefore, an alternative hypothesis to explain the activity of the L35K-H269V variant is that the Arm interacts with AS2, but this interaction does not involve specific contacts between residues L35 and H269. Rather, loss of the bulky imidazole ring at position 269 could alter the local structure of the AS2 region, resulting in loss of multiple inhibitory contacts with the Arm. The long hydrophilic side chain of the L35K substitution could replace contacts typically made with the imidazole ring of the histidine at position 269, restoring local structure and therefore inhibitory interdomain contacts of AS2 with the Arm. Regardless, this model is very similar to the model for the contacts between the Arm of AraC and the region of its DBD that aligns with RhaR AS2 (68).

In the absence of effector, AraC Arm residues L9 and L10 are predicted to make hydrophobic contacts with residues L237 and L238 in AS2 to inhibit transcription

activation (68). Additionally, AraC residue R251 in AS2 is predicted to make an electrostatic interaction with either E3 or D7 in the AraC Arm (68). Although our mutagenesis results do not support a role for the RhaR L37 and L38 (residues that align with AraC L9 and L10) in inhibitory contacts in the absence of effector, RhaR residue L35 could make effector-dependent hydrophobic contacts with the DBD at AS2. In RhaR, we predict that L35 could contact residue Y265 and/or L266 in the AS2 region. RhaR residues Y265 and L266 align with AraC residues L237 and L238 and both are likely to make strong contacts with leucine (167). While we did not identify variants with increased basal activation at either Y265 or L266, a majority of variants at these positions did have decreased basal activity, which could be consistent with a role for these residues in the function of the (-)rhamnose state. RhaR Residue H269 may form an electrostatic interaction with residue D41 in the Arm, similar to the E3 or D7 interaction with R251 in AraC.

It was surprising to us that we identified only a single pair of residues involved in modulation of the RhaR allosteric response. However, RhaR has a lower fold-increase in activity in response to its effector than many other AraC family members, including RhaS (31, 86, 214), potentially supporting the hypothesis that relatively few interdomain contacts may be required for allosteric signaling. This is consistent with findings in RhaS where multiple residues are likely involved in making interdomain contacts for allosteric signaling (Chatterjee, Li, Hunjan and Egan, manuscript in preparation). However, it is plausible that we have not yet identified all of the RhaR interdomain contacts involved in allosteric signaling. We identified a large number of variants at several positions in the RhaR DBD with decreased basal activity. We hypothesize these residues may be involved

in making interdomain contacts and likely are integral to the function of the (-)rhamnose state, although further work is necessary to draw decisive conclusions as to the role of these residues in allosteric signaling. In addition, there may be more regions involved in RhaR interdomain allosteric signaling that we have not yet identified.

Variants with Decreased Basal Activity May Have a Role in Allosteric Signaling. Within this study, 51 of the 99 variants isolated at 20 of 22 positions mutagenized had decreased activity (-) but not (+)rhamnose (Tables 9, 10, 11, 12 and 13). Protein levels for these variants did not correspond to the measured activation levels, indicating decreased protein was not responsible for this phenotype. This led us to hypothesize that decreases in RhaR stability or protein levels may have a greater impact on RhaR (-)rhamnose than (+)rhamnose. Our finding that activity decreased (-) but not (+)rhamnose with declining protein levels support this hypothesis, however the maximum impact attributable solely to stability was determined to have no more than a 3- to 4-fold impact on the activity of the (-) and (+)rhamnose states (Figure 9). However, many of the variants with decreased basal activity had fold defects much greater than this 3-4 fold difference between the (-) and (+)rhamnose states.

For residues critical to maintaining basal activation of RhaR, we would expect a majority of variants at these positions to have a phenotype where activity is decreased (-) but not (+)rhamnose. Our findings indicated to us that these variants did not completely destabilize the protein, but rather specifically impacted basal activation, thereby increasing the allosteric response to L-rhamnose. This led us to hypothesize that these residues likely contribute to the function of RhaR (-)rhamnose and therefore may be involved in allosteric signaling.

Alternatively, although the protein levels tested for representative variants within this class did not correlate with activation levels (-)rhamnose, several of the variants had slightly reduced protein levels in the (-)rhamnose state (Tables 9 and 10). Variants where protein levels were slightly reduced may be in a range where the amount of RhaR binding at *rhaSR* is saturated in the (+)rhamnose state, but not in the (-)rhamnose state. The difference in saturation levels between the (-) and (+)rhamnose states is likely due to the approximate 20-fold difference in DNA binding affinity between the two states (94). Given this difference, in the (-)rhamnose state higher protein levels are required than in the (+)rhamnose state to saturate binding. Therefore, a slight reduction in variant protein levels could explain the phenotype of decreased activity (-) but not (+)rhamnose. However, this does not explain variants where protein levels were wild type but also had the phenotype of decreased activity (-) but not (+)rhamnose. We therefore conclude that residues where a majority of variants fall into this class likely have a role in allosteric signaling in RhaR.

The hypothesis that different protein levels are necessary for saturation of the RhaR binding site at *rhaSR* (-) and (+)rhamnose could also explain the finding that differences in CRP co-activation (-) and (+)rhamnose are eliminated with overexpression of RhaR. At limiting concentrations of RhaR, the RhaR binding site is likely occupied at a lower frequency (-)rhamnose than (+)rhamnose. For maximal activation of the *rhaSR* operon, both RhaR and CRP are required in addition to the α-CTD of RNA polymerase (31, 95). Therefore, we hypothesize that increased CRP co-activation is a result of greater occupancy by RhaR of its binding site in the L-rhamnose-bound state.

C-HTH1 and C-HTH2 Undergo L-Rhamnose-Dependent Conformational Changes. I also investigated the role in allosteric signaling of residues outside of AS2 that align with ToxT residues that are positioned to make interdomain contacts and are located towards the C-terminal ends of each of the two RhaR HTH motifs (79). Mutagenesis of residues in C-HTH2 identified a position, RhaR R295, that may have a role in differentially binding to DNA and possibly also the RhaR NTD in the (-) versus (+)rhamnose states. The following findings support the hypothesis that RhaR R295 may make DNA backbone contacts (-)rhamnose: the wild-type residue is an arginine (the residue that most frequently contacts DNA (165)); multiple substitutions resulted in decreased activity (-)rhamnose; and 88% of a set of ~200 AraC family protein sequences have a residue likely to contact the DNA backbone (R, K, S, T, N or Q) at the aligned position. In contrast, the increased activation of our variants at R295 in the presence of rhamnose implied that this residue has a negative impact on RhaR activity (+)rhamnose. It is possible that the conformation of RhaR (+)rhamnose is such that R295 interferes with DNA binding. Alternatively, RhaR R295 may clash with the NTD in the (+)rhamnose conformation. Although the aligned ToxT residue is not positioned to contact its NTD, it is just outside the distance range for a contact, thus the longer side chain of RhaR R295 (relative to ToxT

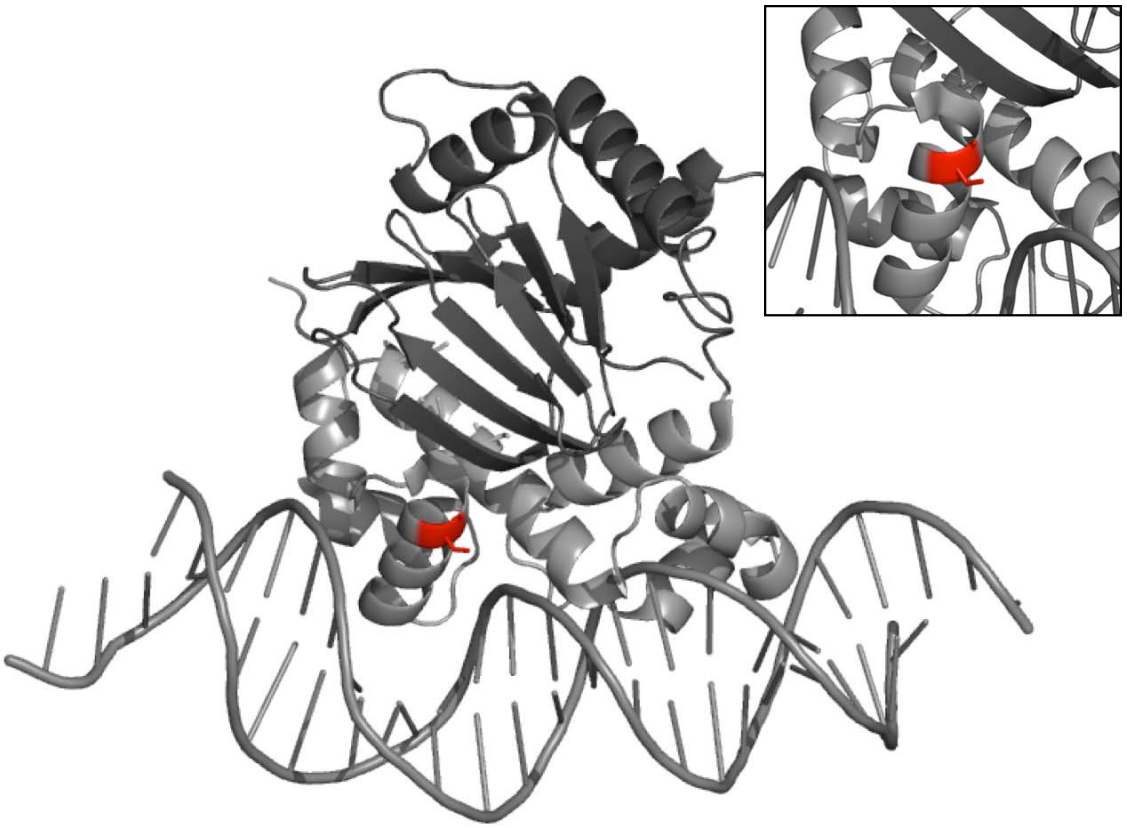


Figure 31. Model of RhaR R295. ToxT crystal structure (79) modeled onto DNA from the MarA co-crystal (43). The NTD is colored dark gray, the DBD is colored light gray. This is equivalent to the "back" view of the model as in Figure 2. The aligned residue in ToxT, S257 (red sticks), is in a position where with a longer side-chain (such as the arginine at RhaR 295) or a slight conformational change this residue could make interdomain contacts with the NTD. Inset, zoomed in view of the side chain at ToxT S257.

S257) or an effector-dependent conformational change might place this RhaR residue in position to contact the NTD (Figure 31).

Mutagenesis of residues in C-HTH1 identified another variant, at position Q246, with differential phenotypes (-) versus (+)rhamnose, indicating that this region also undergoes conformational changes in response to L-rhamnose. The aligned ToxT R209 residue is positioned to make interdomain contacts with the NTD (79) (Figure 32). In addition, there is substantial evidence that AraC residue C183 makes contacts with AraC NTD (+)arabinose (74), and AraC C183 is positioned fairly near the AraC residue that aligns with RhaR Q246 (AraC Q218) in the three-dimensional structure of the AraC DBD (72), and therefore might be positioned to also contact the NTD (Figure 33). We therefore propose that RhaR Q246R is a gain-of-residue-function variant that contacts the NTD (+)rhamnose and increases RhaR activity.

My studies have identified two positions, RhaR Q246 and R295, with variant phenotypes that support the presence of conformational changes toward the C-terminal ends of the two RhaR DNA recognition helices in the (-) versus (+)rhamnose states. The recognition helices of the two HTH motifs in the DNA-bound MarA structure (43) are oriented nearly parallel to one another. In contrast, the recognition helices of both ToxT (in a non-activating state) and AraC-DBD (in the absence of the NTD, not bound to DNA) are not parallel to one another, and thus may undergo a conformational change to enable DNA binding (72, 79). These findings support our model that there is a conformational change in the RhaR DNA recognition helices upon addition of rhamnose, as indicated by the variants at Q246 and R295.

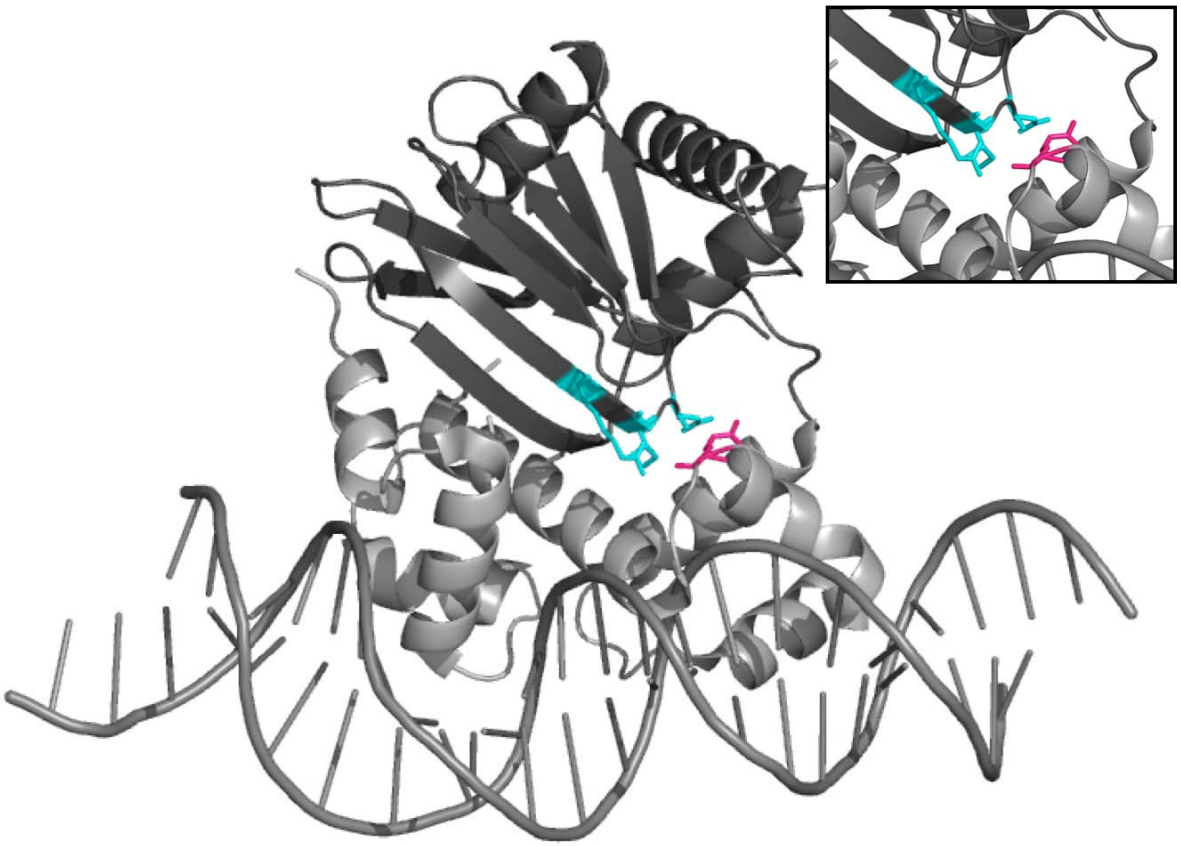


Figure 32. Model of Potential RhaR 246 Interdomain Interactions. ToxT crystal structure (79) modeled onto DNA from the MarA co-crystal (43). NTD shown in dark gray, DBD in light gray. This is equivalent to the "back" view of the model as in Figure 2. The aligned residue in ToxT, R209 (pink sticks), is in a position where it could make interdomain contacts with residues in the NTD (cyan sticks). Inset, zoomed in view of the side chain at ToxT R209 and potential interactions with the NTD.

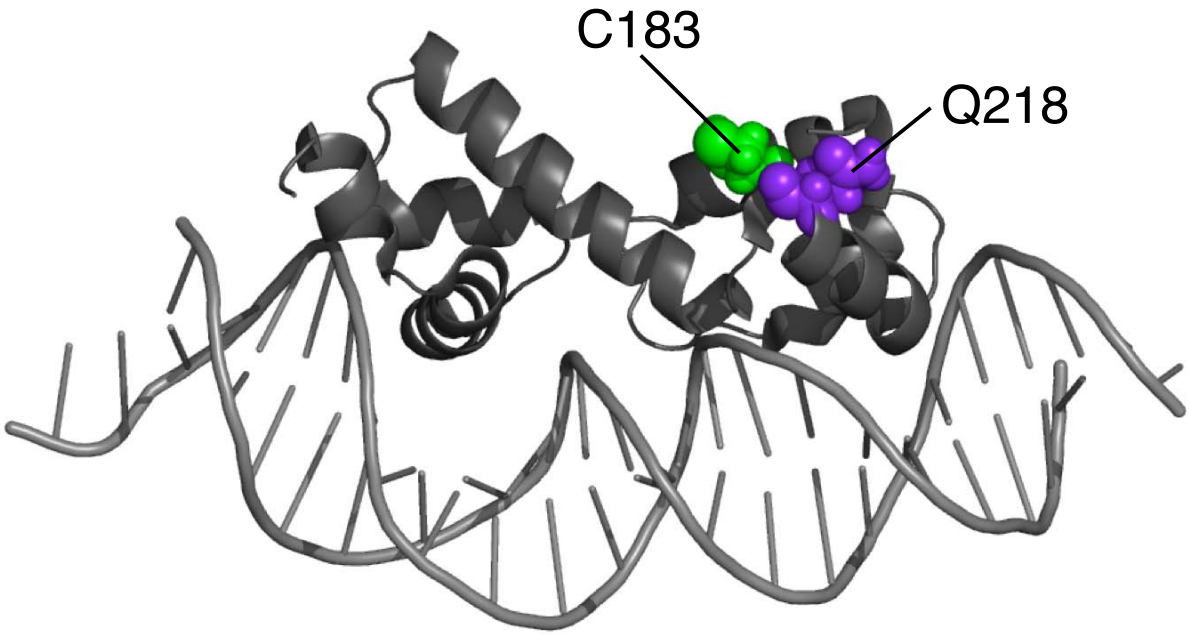


Figure 33. Modeling of RhaR 246 on the AraC DBD. The AraC DBD solution structure (72) modeled onto DNA from the MarA co-crystal (43). The aligned residue in AraC, Q218 (purple spheres) is positioned near AraC reside C183 (green spheres), which likely contacts the AraC NTD (+)arabinose.

We also note that the phenotypes of the RhaR R295 variants support a role for this residue (increasing activity (-)rhamnose and decreasing activity (+)rhamnose) that is opposite to the overall rhamnose response (which increases activity (+)rhamnose). This suggests that while the net effect of the rhamnose-induced conformational changes is an increase in activity (+)rhamnose relative to (-)rhamnose, some individual interdomain interactions may have the opposite effect.

Basal Levels of Active RhaR, but not RhaS, are Sufficient for Transcription Activation. The only known role for RhaR is to increase the concentration of the RhaS protein through activation of the *rhaSR* operon in response to L-rhamnose. Increased concentrations of RhaS then allow RhaS to directly activate transcription of the genes that encode the L-rhamnose catabolic enzymes and L-rhamnose transporter protein (25, 86, 88, 90, 92, 94). Studies on the mRNA induction kinetics of the *rhaSR* and *rhaBAD* promoters indicated a lag in accumulation of *rhaBAD* mRNA, but not of *rhaSR* mRNA after induction with L-rhamnose (86). This finding indicated that, despite being encoded in the same operon, upon addition of L-rhamnose basal levels of active RhaS are insufficient to saturate the *rhaBAD* promoter while basal levels of RhaR are sufficient for saturation of the *rhaSR* promoter (86). Therefore, if a two-step induction by RhaR and RhaS were not utilized by the L-rhamnose regulon, it is likely that basal levels of RhaS would be unable to respond to L-rhamnose in a manner sufficient for utilization of the available sugar.

The RhaR N-terminal Extension Allows Translational Coupling of *rhaS* and *rhaR*. The *rhaR* gene slightly overlaps the *rhaS* gene within the *rhaSR* operon, which is transcribed as a single polycistronic mRNA (25). The RhaS and RhaR proteins consist of 278 and 312 amino acids, respectively (25). Thirty-three of the additional amino acids

Figure 34. Alignment of RhaR and RhaS. Amino acid sequences of the RhaR and RhaS N-terminal regions were aligned using the ClustalW2 program on the EMBL-EBI server. RhaR has an additional 33 amino acids at its N-terminus not present in RhaR. The red dotted line indicates the site where the RhaR Extension is cleaved.

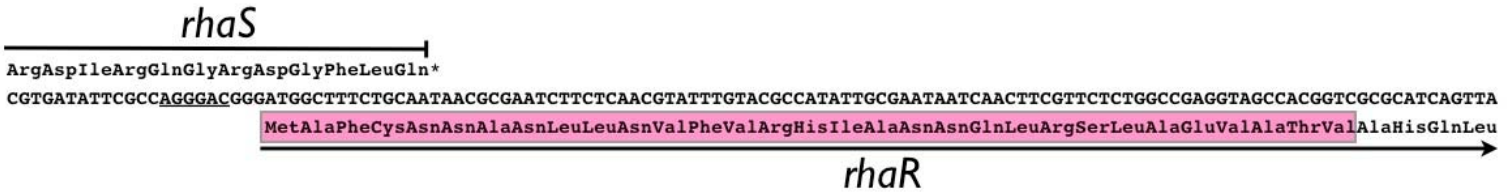


Figure 35. Overlap of the *rhaS* and *rhaR* open reading frames. Sequence of the overlapping region of the *rhaS* and *rhaR* genes. The DNA sequence of this region is shown in between the amino acid sequences for RhaS and RhaR. The amino acid sequence of the RhaS protein is shown above the gene sequence. * indicates the stop codon for *rhaS*. The amino acid sequence of the RhaR protein is shown below the DNA sequence with the RBS underlined. The pink box indicates the N-terminal Extension of RhaR (residues 1-31) that are cleaved. The overlap of the *rhaS* and *rhaR* genes results in a +1 frame shift for the *rhaR* open reading frame.

present in RhaR, compared to RhaS, are located in the N-terminal Extension (144) (Figure 34). Previous results (144), as well as results from this study indicated that the Extension likely does not have a role in allosteric signaling or transcription activation. This is consistent with the recent finding that the first thirty-one amino acids are cleaved and only the cleaved protein is active (Li and Egan, unpublished results). Since the Extension is removed from the protein and is not required for activation, this then begs the question as to the purpose of the Extension.

One clue as to the purpose of the Extension may lie in the coding sequence of the *rhaSR* operon. The overlap of the *rhaR* gene with *rhaS* is generated by the presence of the Extension (Figure 35). This overlap likely results in translational coupling, which is commonly used to coordinate gene expression (215). While the *rhaS* ribosome binding site (RBS) matches the consensus binding sequence (AGGAGG (216)), the RBS for *rhaR* is considerably weaker (GCCAGG (25)). The weaker RBS for *rhaR* is consistent with the finding that expression of the downstream-coupled gene is often dependent on translation of the upstream gene due to a weaker and/or occluded RBS due to formation of stem-loop structures in the mRNA (217-220). Translational coupling of RhaS and RhaR would allow for more rapid accumulation of the RhaS protein, as increased levels of RhaR would result in higher transcription of *rhaSR*. The necessity for rapid accumulation of RhaS becomes apparent considering the slow induction of *rhaBAD* upon the addition of L-rhamnose (86). Accumulation of RhaS, therefore, would result in increased accumulation of the gene products of the *rhaBAD* and *rhaT* operons.

Bicarbonate negatively regulates activation by VirF. Following reports identifying small chemical effectors of the AraC family virulence regulators RegA (145) and

ToxT (79, 80, 146, 186, 221), I screened for effectors of VirF and the closely related Rns protein (170). Structural similarity of the individual ToxT domains with the cognate domains of AraC, in addition to the presence of an effector bound in the effector-binding pocket in the crystal structure of ToxT (79), suggest that perhaps the effector-binding function of the AraC NTD may be conserved within some of the AraC family virulence regulators. Additionally, *in silico* structural modeling of VirF predicts structural similarity of VirF with ToxT and both domains of AraC (Figure 17). Screens for potential effectors of VirF identified bicarbonate as a putative inhibitor of VirF at physiological levels (40mM), reducing VirF-dependent activation by approximately three-fold (Table 27).

Expression of *virF* is regulated in a temperature and pH-dependent manner by the nucleoid associated protein, H-NS (104). Expression of *virF* is permissive at 37°C, however there is 2-fold lower expression at pH 6.0 than at pH 7.0 even at a permissive temperature (104). Ingestion of the organism by the host results in a shift to a temperature that is permissive for the expression of *virF*, however, the acidic environment of the stomach likely results in some pH-dependent repression by H-NS of the *virF* promoter. Transition from the stomach to the small intestine results in an increased pH due to the secretion of bicarbonate from the pancreas (204) and therefore fully permissive conditions for expression of VirF. As a result of increased expression of VirF, without an additional regulatory component expression of the target genes of the VirF protein would also be expressed while the organism is still in the small intestine. Our findings that bicarbonate reduces activation by the VirF protein led me to hypothesize that bicarbonate provides spatial regulation of expression of the T3SS by repressing VirF-dependent activation during passage through the small intestine. Upon reaching the large intestine where infection

Figure 36. Sequence alignment of VirF, ToxT and AraC NTD's. Amino acid sequences of the VirF, AraC and ToxT NTD's were aligned using the ClustalW2 program on the EMBL-EBI server. The region predicted to make protein-protein contacts is indicated by a bar above the sequences with specific residues in ToxT and AraC highlighted in gray boxes. Every tenth residue (relative to the ToxT sequence) is underlined. An (*) indicates identity, (:) indicates strong similarity, (.) indicates weak similarity.

occurs, reduced bicarbonate concentrations result in increased activation by VirF, and thus expression of the genes encoding the T3SS. Future studies are needed to determine if the effect of bicarbonate on VirF is due to a pH-induced change in the protein or if bicarbonate directly binds to VirF as an effector. It would also be interesting to determine what impact, if any, negative regulation by bicarbonate has on DNA binding and dimerization by VirF.

VirF likely forms dimers in solution. DNase footprinting and *in vivo* deletion analysis studies have found that VirF binds to a large region, approximately 100 base pairs in length, upstream of the transcription start point at *virB* (106, 208). The size of this region indicated the presence of more than one VirF binding site, suggesting that VirF binds DNA in an oligomeric state, likely as a dimer (106, 208). Identification of dominant negative alleles of VirF indicated protein-protein interactions (172), providing further support for dimerization of VirF. Lastly, the VirF homolog from *Yersinia enterocolitica* binds to sites at the *yopE, yopH* and *virC* promoters as a dimer (222, 223). Nonetheless, biochemical analysis of the dimeric state of VirF from *S. flexneri* has not been reported.

I therefore investigated the oligomeric state of MBP-VirF^{co} *in vitro*. A preliminary study using gel filtration chromatography of the fusion protein indicated the presence of dimers in solution (Figure 27). Additionally, electrophoretic mobility assays (EMSAs) indicated that both predicted half sites at the *virB* promoter are required for VirF binding. However, only a single binding site is predicted at the *cfaA* promoter where VirF does bind (Figures 25 and 26). More studies are necessary to clarify the oligomeric state of VirF and to better understand the biological role of VirF oligomerization.

Dimerization plays a key role in formation of a competent transcription complex in other AraC family proteins (221, 224-228). To gain insight into a possible dimerization

mechanism for VirF, I looked at alignments of VirF with AraC family proteins with defined dimerization interfaces. In AraC, three leucine residues critical for dimerization (229) anchor the dimer interface (71). This is also the case for of UreR (226) and XylS (227), where the lysines are conserved. None of the three lysines conserved in AraC, UreR and XylS are conserved in VirF (Figure 34). In the case of ToxT, six residues have either been predicted or shown to have a role in dimerization (79, 221). Again, none of these residues are conserved in VirF (Figure 36). The lack of conservation of residues at positions aligning with residues involved in dimerization in AraC and ToxT does not diminish the likelihood of VirF dimerization. Mutagenesis of this region will likely provide valuable information regarding the specific residues involved in VirF dimer formation, which could be confirmed via a variety of techniques such as bacterial two-hybrid analysis or dominant-negative studies in addition to biochemical analysis of the dimeric state of variants via gel filtration chromatography.

SE-1 inhibits DNA binding by RhaR and VirF. The near ubiquity of AraC family regulators within sequenced bacterial genomes (24) (many of are required for bacterial virulence (29, 51, 230-232)) makes this large family of bacterial transcription regulators a good target for development of novel anti-microbial drugs. The small molecule inhibitor, SE-1, inhibits DNA binding by the RhaS, RhaR (148) and VirF proteins (147). I demonstrated, using EMSAs, that SE-1 inhibits *in vitro* DNA binding for RhaR (148) and provided preliminary evidence that it also inhibits VirF. These results, taken together with the remainder of our findings, indicate that the mode of action of SE-1 likely involves inhibition of DNA binding by the conserved DBD of the AraC family (147, 148, 157). Specificity of the inhibitor for the conserved domain may result in inhibition of multiple

members of the AraC family, making SE-1 a good candidate for further development as an anti-microbial drug. Furthermore, many AraC family virulence regulators, including VirF, are non-essential for bacterial survival (58, 59, 147, 157). The non-essential nature of these regulators is yet another property that makes AraC family regulators favorable targets for development of novel anti-infective agents, as targeting these regulators is hypothesized to result in reduced selective pressure to develop resistance compared with the current antibiotic treatments available (60-63, 147, 157). Although other inhibitors of AraC family regulators have been identified (233-238), the large size and diversity within this family of proteins is expected to limit the coverage of potential inhibitors. This, combined with the implicit uncertainty associated with drug development, highlights the need for identification of multiple inhibitors of the AraC family to increase the probability of successful development of anti-infective agents targeting AraC family regulators.

Regulatory interdomain contacts are likely conserved within some members of the AraC family. Overall, the studies presented in this work contribute to the general understanding of mechanisms of transcription activation used by the AraC family of activators. This work supports a possible mechanism where interdomain contacts inhibit transcription activation by RhaR in the absence of effector and sheds light on the basic requirements for activation by VirF. Although it is not yet clear if VirF also responds to an effector to regulate transcription activation, it is likely that the mechanism of transcription activation may be conserved with that of RhaR and other members of the AraC family such as AraC and ToxT.

In this work, we found that the AS2 region of RhaR is likely involved in making interdomain contacts that inhibit transcription activation in the absence of rhamnose.

Residues in this same AS2 region of AraC and ToxT have also been found to have a role in making inhibitory interdomain contacts (43, 68, 79, 161). Furthermore, the AS2 region of the Rns protein has recently been found to have a role in the transcriptional activation ability of the protein, likely through interdomain contacts with the Rns NTD (Koppolu, Munson and Egan, manuscript in preparation). These findings lead us to hypothesize that the AS2 region of the AraC family DBD may have a conserved role in regulation of transcription activation for at least some of the members of the family with similarity to AraC over its full-length, including VirF.

Unlike AraC (64-68) and RhaR (144), sequence alignments and I-TASSER structural models (159, 160) do not predict that VirF utilizes an N-terminal Arm that could make interdomain contacts with the DBD at AS2. However, VirF may utilize an alternative region at the N-terminus of the NTD, which we call regulatory site 2 (RS2). The ToxT protein does not utilize an N-terminal Arm, rather residues in the first two β strands (β1 and β2), as well as the effector, make interdomain contacts with the AS2 region of ToxT (79). In Rns, the residues in RS2 that align with residues in β2 of ToxT have also been found to have a role in transcription activation and are predicted to make interdomain contacts with the AS2 region (Koppolu, Munson and Egan, manuscript in preparation). Due to the high similarity of VirF with Rns and ToxT (Table 20), we hypothesize that the RS2 region of VirF may also have a role in making interdomain contacts involved in regulation of transcription activation.

Given the parallels found between AraC, ToxT, RhaR and Rns, we predict a conserved model where the AS2 region of the DBD is contacted by either the Arm or RS2 to regulate transcription activation. The result of contacts at AS2 could be inhibitory, as found

in RhaR, AraC (68) and ToxT (79), or stimulatory as found for the RhaS and Rns proteins (Chatterjee, Li, Hunjan and Egan, manuscript in preparation and Koppolu, Munson and Egan, manuscript in preparation). Regardless of the outcome of these contacts (inhibitory or stimulatory), there is growing evidence supporting a conservation of the interdomain contacts between the NTD and the AS2 region, which likely result in regulation of transcription activation by various members of the AraC family.

Chapter VI: References

1. **Snyder LaC, Wendy.** 2007. Molecular Genetics of Bacteria 3rd. ed. ASM Press, Washiington DC.
2. **Lee DJ, Minchin SD, Busby SJ.** 2012. Activating transcription in bacteria. Annu Rev Microbiol. **66:**125-152.
3. **Burgess RR.** 1969. Separation and characterization of the subunits of ribonucleic acid polymerase. J Biol Chem. **244:**6168-6176.
4. **Burgess RR.** 1971. RNA polymerase. Annu Rev Biochem. **40:**711-740.
5. **Helmann JD, Chamberlin MJ.** 1988. Structure and function of bacterial sigma factors. Annu Rev Biochem. **57:**839-872.
6. **Gross CA, Chan C, Dombroski A, Gruber T, Sharp M, Tupy J, Young B.** 1998. The functional and regulatory roles of sigma factors in transcription. Cold Spring Harbor symposia on quantitative biology. **63:**141-155.
7. **Ishihama A.** 2000. Functional modulation of *Escherichia coli* RNA polymerase. Annu Rev Microbiol. **54:**499-518.
8. **Gross CA, Lonetto, M., Losick, R.** 1992. Bacterial Sigma Factors, p. 129-176 *In* McKnight SLaY, K.R. (ed.), Transcriptional Regulation. Cold Spring Harbour Press, New York.
9. **Dombroski AJ, Walter WA, Record MT, Jr., Siegele DA, Gross CA.** 1992. Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA. Cell. **70:**501-512.
10. **Paget MS, Helmann JD.** 2003. The sigma70 family of sigma factors. Genome biology. **4:**203.

11. **Hawley DK, McClure WR.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11:**2237-2255.

12. **Mecsas J, Cowing DW, Gross CA.** 1991. Development of RNA polymerase-promoter contacts during open complex formation. J Mol Biol. **220:**585-597.

13. **Spassky A, Kirkegaard K, Buc H.** 1985. Changes in the DNA structure of the *lac*UV5 promoter during formation of an open complex with *Escherichia coli* RNA polymerase. Biochemistry. **24:**2723-2731.

14. **Schickor P, Metzger W, Werel W, Lederer H, Heumann H.** 1990. Topography of intermediates in transcription initiation of *E.coli*. EMBO J. **9:**2215-2220.

15. **Browning DF, Busby SJ.** 2004. The regulation of bacterial transcription initiation. Nat Rev Microbiol. **2:**57-65.

16. **Busby S, Ebright RH.** 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell. **79:**743-746.

17. **Rhodius VA, Busby SJ.** 1998. Positive activation of gene expression. Curr Opin Microbiol. **1:**152-159.

18. **Dove SL, Darst SA, Hochschild A.** 2003. Region 4 of sigma as a target for transcription regulation. Mol Microbiol. **48:**863-874.

19. **Busby S, Ebright RH.** 1997. Transcription activation at class II CAP-dependent promoters. Mol Microbiol. **23:**853-859.

20. **Brown NL, Stoyanov JV, Kidd SP, Hobman JL.** 2003. The MerR family of transcriptional regulators. FEMS Microbiol Rev. **27:**145-163.

21. **Sheridan SD, Benham CJ, Hatfield GW.** 1998. Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced

DNA duplex destabilization in an upstream activating sequence. J Biol Chem. **273:**21298-21308.

22. **Lawson CL, Swigon D, Murakami KS, Darst SA, Berman HM, Ebright RH.** 2004. Catabolite activator protein: DNA binding and transcription activation. Curr Opin Struct Biol. **14:**10-20.

23. **Busby S, Ebright RH.** 1999. Transcription activation by catabolite activator protein (CAP). J Mol Biol. **293:**199-213.

24. **Ibarra JA, Perez-Rueda E, Segovia L, Puente JL.** 2008. The DNA-binding domain as a functional indicator: the case of the AraC/XylS family of transcription factors. Genetica. **133:**65-76.

25. **Tobin JF, Schleif RF.** 1987. Positive regulation of the *Escherichia coli* L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes. J Mol Biol. **196:**789-799.

26. **Gallegos MT, Michan C, Ramos JL.** 1993. The XylS/AraC family of regulators. Nucleic Acids Res. **21:**807-810.

27. **Ramos JL, Rojo F, Zhou L, Timmis KN.** 1990. A family of positive regulators related to the *Pseudomonas putida* TOL plasmid XylS and the *Escherichia coli* AraC activators. Nucleic Acids Res. **18:**2149-2152.

28. **Egan SM.** 2002. Growing repertoire of AraC/XylS activators. J Bacteriol. **184:**5529-5532.

29. **Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL.** 1997. AraC/XylS family of transcriptional regulators. Microbiol Mol Biol Rev. **61:**393-410.

30. **Dangi B, Gronenborn AM, Rosner JL, Martin RG.** 2004. Versatility of the carboxy-terminal domain of the alpha subunit of RNA polymerase in transcriptional activation: use of the DNA contact site as a protein contact site for MarA. Mol Microbiol. **54:**45-59.

31. **Holcroft CC, Egan SM.** 2000. Interdependence of activation at *rhaSR* by cyclic AMP receptor protein, the RNA polymerase alpha subunit C-terminal domain, and RhaR. J Bacteriol. **182:**6774-6782.

32. **Jair KW, Martin RG, Rosner JL, Fujita N, Ishihama A, Wolf RE, Jr.** 1995. Purification and regulatory properties of MarA protein, a transcriptional activator of *Escherichia coli* multiple antibiotic and superoxide resistance promoters. J Bacteriol. **177:**7100-7104.

33. **Jair KW, Yu X, Skarstad K, Thony B, Fujita N, Ishihama A, Wolf RE, Jr.** 1996. Transcriptional activation of promoters of the superoxide and multiple antibiotic resistance regulons by Rob, a binding protein of the *Escherichia coli* origin of chromosomal replication. J Bacteriol. **178:**2507-2513.

34. **Landini P, Gaal T, Ross W, Volkert MR.** 1997. The RNA polymerase alpha subunit carboxyl-terminal domain is required for both basal and activated transcription from the *alkA* promoter. J Biol Chem. **272:**15914-15919.

35. **Ruiz R, Ramos JL, Egan SM.** 2001. Interactions of the XylS regulators with the C-terminal domain of the RNA polymerase alpha subunit influence the expression level from the cognate *Pm* promoter. FEBS Lett. **491:**207-211.

36. **Shah IM, Wolf RE, Jr.** 2004. Novel protein--protein interaction between *Escherichia coli* SoxS and the DNA binding determinant of the RNA polymerase alpha subunit:

SoxS functions as a co-sigma factor and redeploys RNA polymerase from UP-element-containing promoters to SoxS-dependent promoters during oxidative stress. J Mol Biol. **343:**513-532.

37. **Wickstrum JR, Egan SM.** 2004. Amino acid contacts between sigma 70 domain 4 and the transcription activators RhaS and RhaR. J Bacteriol. **186:**6277-6285.

38. **Bhende PM, Egan SM.** 2000. Genetic evidence that transcription activation by RhaS involves specific amino acid contacts with sigma 70. J Bacteriol. **182:**4959-4969.

39. **Grainger DC, Belyaeva TA, Lee DJ, Hyde EI, Busby SJ.** 2004. Transcription activation at the *Escherichia coli melAB* promoter: interactions of MelR with the C-terminal domain of the RNA polymerase alpha subunit. Mol Microbiol. **51:**1311-1320.

40. **Grainger DC, Webster CL, Belyaeva TA, Hyde EI, Busby SJ.** 2004. Transcription activation at the *Escherichia coli melAB* promoter: interactions of MelR with its DNA target site and with domain 4 of the RNA polymerase sigma subunit. Mol Microbiol. **51:**1297-1309.

41. **Landini P, Busby SJ.** 1999. The *Escherichia coli* Ada protein can interact with two distinct determinants in the sigma70 subunit of RNA polymerase according to promoter architecture: identification of the target of Ada activation at the *alkA* promoter. J Bacteriol. **181:**1524-1529.

42. **Schüller A, Slater AW, Norambuena T, Cifuentes JJ, Almonacid LI, Melo F.** 2012. Computer-based annotation of putative AraC/XylS-family transcription factors of known structure but unknown function. J. Biomed. Biotechnol. **2012:**103132.

43. **Rhee S, Martin RG, Rosner JL, Davies DR.** 1998. 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. **95:**10413-10418.

44. **Spooner RA, Lindsay K, Franklin FC.** 1986. Genetic, functional and sequence analysis of the *xylR* and *xylS* regulatory genes of the TOL plasmid pWW0. J Gen Microbiol. **132:**1347-1358.

45. **Webster C, Gardner L, Busby S.** 1989. The *Escherichia coli melR* gene encodes a DNA-binding protein with affinity for specific sequences located in the melibiose-operon regulatory region. Gene. **83:**207-213.

46. **Demple B, Sedgwick B, Robins P, Totty N, Waterfield MD, Lindahl T.** 1985. Active site and complete sequence of the suicidal methyltransferase that counters alkylation mutagenesis. Proceedings of the National Academy of Sciences of the United States of America. **82:**2688-2692.

47. **Hakura A, Morimoto K, Sofuni T, Nohmi T.** 1991. Cloning and characterization of the *Salmonella typhimurium ada* gene, which encodes O6-methylguanine-DNA methyltransferase. J Bacteriol. **173:**3663-3672.

48. **Morohoshi F, Hayashi K, Munakata N.** 1990. *Bacillus subtilis ada* operon encodes two DNA alkyltransferases. Nucleic Acids Res. **18:**5473-5480.

49. **Amabile-Cuevas CF, Demple B.** 1991. Molecular characterization of the *soxRS* genes of *Escherichia coli*: two genes control a superoxide stress regulon. Nucleic Acids Res. **19:**4479-4484.

50. **Cohen SP, Hachler H, Levy SB.** 1993. Genetic and functional analysis of the multiple antibiotic resistance (*mar*) locus in *Escherichia coli*. J Bacteriol. **175:**1484-1492.

51. **Champion GA, Neely MN, Brennan MA, DiRita VJ.** 1997. A branch in the ToxR regulatory cascade of *Vibrio cholerae* revealed by characterization of *toxT* mutant strains. Mol Microbiol. **23:**323-331.

52. **DiRita VJ, Parsot C, Jander G, Mekalanos JJ.** 1991. Regulatory cascade controls virulence in *Vibrio cholerae*. Proceedings of the National Academy of Sciences of the United States of America. **88:**5403-5407.

53. **Higgins DE, Nazareno E, DiRita VJ.** 1992. The virulence gene activator ToxT from *Vibrio cholerae* is a member of the AraC family of transcriptional activators. J Bacteriol. **174:**6974-6980.

54. **Brown RC, Taylor RK.** 1995. Organization of *tcp*, *acf*, and *toxT* genes within a ToxT-dependent operon. Mol Microbiol. **16:**425-439.

55. **Yu RR, DiRita VJ.** 1999. Analysis of an autoregulatory loop controlling ToxT, cholera toxin, and toxin-coregulated pilus production in *Vibrio cholerae*. J Bacteriol. **181:**2584-2592.

56. **Caron J, Coffield LM, Scott JR.** 1989. A plasmid-encoded regulatory gene, *rns*, required for expression of the CS1 and CS2 adhesins of enterotoxigenic *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America. **86:**963-967.

57. **Adler B, Sasakawa C, Tobe T, Makino S, Komatsu K, Yoshikawa M.** 1989. A dual transcriptional activation system for the 230 kb plasmid genes coding for virulence-associated antigens of *Shigella flexneri*. Mol Microbiol. **3:**627-635.

58. **Casaz P, Garrity-Ryan LK, McKenney D, Jackson C, Levy SB, Tanaka SK, Alekshun MN.** 2006. MarA, SoxS and Rob function as virulence factors in an

Escherichia coli murine model of ascending pyelonephritis. Microbiology. **152:**3643-3650.

59. **Flashner Y, Mamroud E, Tidhar A, Ber R, Aftalion M, Gur D, Lazar S, Zvi A, Bino T, Ariel N, Velan B, Shafferman A, Cohen S.** 2004. Generation of *Yersinia pestis* attenuated strains by signature-tagged mutagenesis in search of novel vaccine candidates. Infect. Immun. **72:**908-915.

60. **Rasko DA, Sperandio V.** 2010. Anti-virulence strategies to combat bacteria-mediated disease. Nat Rev Drug Discov. **9:**117-128.

61. **Hughes D.** 2003. Exploiting genomics, genetics and chemistry to combat antibiotic resistance. Nat Rev Genet. **4:**432-441.

62. **Knowles DJ.** 1997. New strategies for antibacterial drug design. Trends Microbiol. **5:**379-383.

63. **Schmidt FR.** 2004. The challenge of multidrug resistance: actual strategies in the development of novel antibacterials. Appl Microbiol Biotechnol. **63:**335-343.

64. **Saviola B, Seabold R, Schleif RF.** 1998. Arm-domain interactions in AraC. J Mol Biol. **278:**539-548.

65. **Reed WL, Schleif RF.** 1999. Hemiplegic mutations in AraC protein. J Mol Biol. **294:**417-425.

66. **Harmer T, Wu M, Schleif R.** 2001. The role of rigidity in DNA looping-unlooping by AraC. Proceedings of the National Academy of Sciences of the United States of America. **98:**427-431.

67. **Wu M, Schleif R.** 2001. Strengthened arm-dimerization domain interactions in AraC. J Biol Chem. **276:**2562-2564.

68. **Wu M, Schleif R.** 2001. Mapping arm-DNA-binding domain interactions in AraC. J Mol Biol. **307:**1001-1009.

69. **Schleif R.** 2000. Regulation of the L-arabinose operon of *Escherichia coli*. Trends Genet. **16:**559-565.

70. **Weldon JE, Rodgers ME, Larkin C, Schleif RF.** 2007. Structure and properties of a truely apo form of AraC dimerization domain. Proteins. **66:**646-654.

71. **Soisson SM, MacDougall-Shackleton B, Schleif R, Wolberger C.** 1997. Structural basis for ligand-regulated oligomerization of AraC. Science. **276:**421-425.

72. **Rodgers ME, Schleif R.** 2009. Solution structure of the DNA binding domain of AraC protein. Proteins. **77:**202-208.

73. **Seabold RR, Schleif RF.** 1998. Apo-AraC actively seeks to loop. J Mol Biol. **278:**529-538.

74. **Cole SD, Schleif R.** 2012. A new and unexpected domain-domain interaction in the AraC protein. Proteins. **80:**1465-1475.

75. **Lobell RB, Schleif RF.** 1990. DNA looping and unlooping by AraC protein. Science. **250:**528-532.

76. **Ghosh M, Schleif RF.** 2001. Biophysical evidence of arm-domain interactions in AraC. Anal Biochem. **295:**107-112.

77. **Ross JJ, Gryczynski U, Schleif R.** 2003. Mutational analysis of residue roles in AraC function. J Mol Biol. **328:**85-93.

78. **Dunn TM, Hahn S, Ogden S, Schleif RF.** 1984. An operator at -280 base pairs that is required for repression of *araBAD* operon promoter: addition of DNA helical turns

between the operator and promoter cyclically hinders repression. Proceedings of the National Academy of Sciences of the United States of America. **81:**5017-5020.

79. **Lowden MJ, Skorupski K, Pellegrini M, Chiorazzo MG, Taylor RK, Kull FJ.** 2010. Structure of *Vibrio cholerae* ToxT reveals a mechanism for fatty acid regulation of virulence genes. Proceedings of the National Academy of Sciences of the United States of America. **107:**2860-2865.

80. **Chatterjee A, Dutta PK, Chowdhury R.** 2007. Effect of fatty acids and cholesterol present in bile on expression of virulence factors and motility of *Vibrio cholerae*. Infect. Immun. **75:**1946-1953.

81. **Dominguez-Cuevas P, Marin P, Busby S, Ramos JL, Marques S.** 2008. Roles of effectors in XylS-dependent transcription activation: intramolecular domain derepression and DNA binding. J Bacteriol. **190:**3118-3128.

82. **Dominguez-Cuevas P, Marin P, Marques S, Ramos JL.** 2008. XylS-*Pm* promoter interactions through two helix-turn-helix motifs: identifying XylS residues important for DNA binding and activation. J Mol Biol. **375:**59-69.

83. **Michán C, Zhou L, Gallegos MT, Timmis KN, Ramos JL.** 1992. Identification of critical amino-terminal regions of XylS. The positive regulator encoded by the TOL plasmid. J Biol Chem. **267:**22897-22901.

84. **Michán C, Kessler B, de Lorenzo V, Timmis KN, Ramos JL.** 1992. XylS domain interactions can be deduced from intraallelic dominance in double mutants of *Pseudomonas putida*. Mol Gen Genet. **235:**406-412.

85. **Baldoma L, Badia J, Sweet G, Aguilar J.** 1990. Cloning, mapping and gene product identification of *rhaT* from *Escherichia coli* K12. FEMS Microbiol Lett. **60:**103-107.

86. **Egan SM, Schleif RF.** 1993. A regulatory cascade in the induction of *rhaBAD*. J Mol Biol. **234:**87-98.

87. **Power J.** 1967. The L-rhamnose genetic system in *Escherichia coli* K-12. Genetics. **55:**557-568.

88. **Tate CG, Muiry JA, Henderson PJ.** 1992. Mapping, cloning, expression, and sequencing of the *rhaT* gene, which encodes a novel L-rhamnose-H^+ transport protein in *Salmonella typhimurium* and *Escherichia coli*. J Biol Chem. **267:**6923-6932.

89. **Badia J, Baldoma L, Aguilar J, Boronat A.** 1989. Identification of the *rhaA*, *rhaB* and *rhaD* gene products from *Escherichia coli* K-12. FEMS Microbiol Lett. **53:**253-257.

90. **Tobin JF, Schleif RF.** 1990. Transcription from the *rha* operon p_{sr} promoter. J Mol Biol. **211:**1-4.

91. **Via P, Badia J, Baldoma L, Obradors N, Aguilar J.** 1996. Transcriptional regulation of the *Escherichia coli rhaT* gene. Microbiology. **142 (Pt 7):**1833-1840.

92. **Egan SM, Schleif RF.** 1994. DNA-dependent renaturation of an insoluble DNA binding protein. Identification of the RhaS binding site at *rhaBAD*. J Mol Biol. **243:**821-829.

93. **Wickstrum JR, Skredenske JM, Kolin A, Jin DJ, Fang J, Egan SM.** 2007. Transcription activation by the DNA-binding domain of the AraC family protein RhaS in the absence of its effector-binding domain. J Bacteriol. **189:**4984-4993.

94. **Tobin JF, Schleif RF.** 1990. Purification and properties of RhaR, the positive regulator of the L-rhamnose operons of *Escherichia coli*. J Mol Biol. **211:**75-89.

95. **Wickstrum JR, Santangelo TJ, Egan SM.** 2005. Cyclic AMP receptor protein and RhaR synergistically activate transcription from the L-rhamnose-responsive *rhaSR* promoter in *Escherichia coli*. J Bacteriol. **187:**6708-6718.

96. **Kolin A, Jevtic V, Swint-Kruse L, Egan SM.** 2007. Linker regions of the RhaS and RhaR proteins. J Bacteriol. **189:**269-271.

97. **Sakai T, Sasakawa C, Makino S, Kamata K, Yoshikawa M.** 1986. Molecular cloning of a genetic determinant for Congo red binding ability which is essential for the virulence of *Shigella flexneri*. Infect. Immun. **51:**476-482.

98. **Sasakawa C, Komatsu K, Tobe T, Suzuki T, Yoshikawa M.** 1993. Eight genes in region 5 that form an operon are essential for invasion of epithelial cells by *Shigella flexneri* 2a. J Bacteriol. **175:**2334-2346.

99. **Sansonetti PJ, Kopecko DJ, Formal SB.** 1982. Involvement of a plasmid in the invasive ability of *Shigella flexneri*. Infect. Immun. **35:**852-860.

100. **Blocker A, Jouihri N, Larquet E, Gounon P, Ebel F, Parsot C, Sansonetti P, Allaoui A.** 2001. Structure and composition of the *Shigella flexneri* "needle complex", a part of its type III secreton. Mol Microbiol. **39:**652-663.

101. **High N, Mounier J, Prevost MC, Sansonetti PJ.** 1992. IpaB of *Shigella flexneri* causes entry into epithelial cells and escape from the phagocytic vacuole. EMBO J. **11:**1991-1999.

102. **Zychlinsky A, Kenny B, Menard R, Prevost MC, Holland IB, Sansonetti PJ.** 1994. IpaB mediates macrophage apoptosis induced by *Shigella flexneri*. Mol Microbiol. **11:**619-627.

103. **Falconi M, Colonna B, Prosseda G, Micheli G, Gualerzi CO.** 1998. 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 J. **17:**7033-7043.

104. **Prosseda G, Fradiani PA, Di Lorenzo M, Falconi M, Micheli G, Casalino M, Nicoletti M, Colonna B.** 1998. A role for H-NS in the regulation of the *virF* gene of Shigella and enteroinvasive *Escherichia coli*. Res Microbiol. **149:**15-25.

105. **Prosseda G, Falconi M, Giangrossi M, Gualerzi CO, Micheli G, Colonna B.** 2004. The *virF* promoter in Shigella: more than just a curved DNA stretch. Mol Microbiol. **51:**523-537.

106. **Tobe T, Yoshikawa M, Mizuno T, Sasakawa C.** 1993. Transcriptional control of the invasion regulatory gene *virB* of *Shigella flexneri*: activation by VirF and repression by H-NS. J Bacteriol. **175:**6142-6149.

107. **Sakai T, Sasakawa C, Yoshikawa M.** 1988. Expression of four virulence antigens of *Shigella flexneri* is positively regulated at the transcriptional level by the 30 kiloDalton VirF protein. Mol Microbiol. **2:**589-597.

108. **Tran CN, Giangrossi M, Prosseda G, Brandi A, Di Martino ML, Colonna B, Falconi M.** 2011. A multifactor regulatory circuit involving H-NS, VirF and an antisense RNA modulates transcription of the virulence gene *icsA* of *Shigella flexneri*. Nucleic Acids Res. **39:**8122-8134.

109. **Tobe T, Nagai S, Okada N, Adler B, Yoshikawa M, Sasakawa C.** 1991. Temperature-regulated expression of invasion genes in *Shigella flexneri* is

controlled through the transcriptional activation of the *virB* gene on the large plasmid. Mol Microbiol. **5:**887-893.

110. **Kane KA, Dorman CJ.** 2012. VirB-mediated positive feedback control of the virulence gene regulatory cascade of *Shigella flexneri*. J Bacteriol. **194:**5264-5273.

111. **Niyogi SK.** 2005. Shigellosis. J Microbiol. **43:**133-143.

112. **Prevention USDoHaHSCfDCa.** 2013. Antibiotic Resistance Threats in the United States, 2013. Centers for Disease Control and Prevention.

113. **Bennish ML, Wojtyniak BJ.** 1991. Mortality due to shigellosis: community and hospital data. Rev Infect Dis. **13 Suppl 4:**S245-251.

114. **Washington Winn Jr. SA, William Janda, Elmer Koneman, Gary Procop, Paul Schreckenberger, Gail Woods.** 2006. Color Atlas and Textbook of Diagnostic Microbiology, 6th ed. Lippincott Williams & Wilkins, Philadelphia, PA.

115. **DuPont HL, Levine MM, Hornick RB, Formal SB.** 1989. Inoculum size in shigellosis and implications for expected mode of transmission. J Infect Dis. **159:**1126-1128.

116. **Schroeder GN, Hilbi H.** 2008. Molecular pathogenesis of Shigella spp.: controlling host cell signaling, invasion, and death by type III secretion. Clin Microbiol Rev. **21:**134-156.

117. **Sansonetti PJ, Arondel J, Cantey JR, Prevost MC, Huerre M.** 1996. Infection of rabbit Peyer's patches by *Shigella flexneri*: effect of adhesive or invasive bacterial phenotypes on follicle-associated epithelium. Infect. Immun. **64:**2752-2764.

118. **Wassef JS, Keren DF, Mailloux JL.** 1989. Role of M cells in initial antigen uptake and in ulcer formation in the rabbit intestinal loop model of shigellosis. Infect. Immun. **57:**858-863.

119. **Mounier J, Vasselon T, Hellio R, Lesourd M, Sansonetti PJ.** 1992. *Shigella flexneri* enters human colonic Caco-2 epithelial cells through the basolateral pole. Infect. Immun. **60:**237-248.

120. **Man AL, Prieto-Garcia ME, Nicoletti C.** 2004. Improving M cell mediated transport across mucosal barriers: do certain bacteria hold the keys? Immunology. **113:**15-22.

121. **Islam D, Veress B, Bardhan PK, Lindberg AA, Christensson B.** 1997. In situ characterization of inflammatory responses in the rectal mucosae of patients with shigellosis. Infect. Immun. **65:**739-749.

122. **Zychlinsky A, Prevost MC, Sansonetti PJ.** 1992. *Shigella flexneri* induces apoptosis in infected macrophages. Nature. **358:**167-169.

123. **Zychlinsky A, Thirumalai K, Arondel J, Cantey JR, Aliprantis AO, Sansonetti PJ.** 1996. In vivo apoptosis in *Shigella flexneri* infections. Infect. Immun. **64:**5357-5365.

124. **Sansonetti PJ, Ryter A, Clerc P, Maurelli AT, Mounier J.** 1986. Multiplication of *Shigella flexneri* within HeLa cells: lysis of the phagocytic vacuole and plasmid-mediated contact hemolysis. Infect. Immun. **51:**461-469.

125. **Bernardini ML, Mounier J, d'Hauteville H, Coquis-Rondon M, Sansonetti PJ.** 1989. Identification of *icsA*, a plasmid locus of *Shigella flexneri* that governs bacterial intra- and intercellular spread through interaction with F-actin. Proceedings of the National Academy of Sciences of the United States of America. **86:**3867-3871.

126. **Lett MC, Sasakawa C, Okada N, Sakai T, Makino S, Yamada M, Komatsu K, Yoshikawa M.** 1989. *virG*, a plasmid-coded virulence gene of *Shigella flexneri*: identification of the VirG protein and determination of the complete coding sequence. J Bacteriol. **171:**353-359.

127. **Sansonetti PJ, Arondel J, Fontaine A, d'Hauteville H, Bernardini ML.** 1991. *ompB* (osmo-regulation) and *icsA* (cell-to-cell spread) mutants of *Shigella flexneri*: vaccine candidates and probes to study the pathogenesis of shigellosis. Vaccine. **9:**416-422.

128. **Jennison AV, Verma NK.** 2004. *Shigella flexneri* infection: pathogenesis and vaccine development. FEMS Microbiol Rev. **28:**43-58.

129. **Buchrieser C, Glaser P, Rusniok C, Nedjari H, D'Hauteville H, Kunst F, Sansonetti P, Parsot C.** 2000. The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. Mol Microbiol. **38:**760-771.

130. **Dorman CJ, Porter ME.** 1998. The Shigella virulence gene regulatory cascade: a paradigm of bacterial gene control mechanisms. Mol Microbiol. **29:**677-684.

131. **Le Gall T, Mavris M, Martino MC, Bernardini ML, Denamur E, Parsot C.** 2005. Analysis of virulence plasmid gene expression defines three classes of effectors in the type III secretion system of *Shigella flexneri*. Microbiology. **151:**951-962.

132. **Blocker A, Gounon P, Larquet E, Niebuhr K, Cabiaux V, Parsot C, Sansonetti P.** 1999. The tripartite type III secreton of *Shigella flexneri* inserts IpaB and IpaC into host membranes. J Cell Biol. **147:**683-693.

133. **Ogawa M, Handa Y, Ashida H, Suzuki M, Sasakawa C.** 2008. The versatility of Shigella effectors. Nat Rev Microbiol. **6:**11-16.

134. **Phalipon A, Sansonetti PJ.** 2007. Shigella's ways of manipulating the host intestinal innate and adaptive immune system: a tool box for survival? Immunol Cell Biol. **85:**119-129.

135. **Marteyn B, Gazi A, Sansonetti P.** 2012. *Shigella*: a model of virulence regulation in vivo. Gut microbes. **3:**104-120.

136. **Cornelis GR.** 2006. The type III secretion injectisome. Nat Rev Microbiol. **4:**811-825.

137. **Ghosh P.** 2004. Process of protein transport by the type III secretion system. Microbiol Mol Biol Rev. **68:**771-795.

138. **Menard R, Sansonetti P, Parsot C.** 1994. The secretion of the *Shigella flexneri* Ipa invasins is activated by epithelial cells and controlled by IpaB and IpaD. EMBO J. **13:**5293-5302.

139. **Veenendaal AK, Hodgkinson JL, Schwarzer L, Stabat D, Zenk SF, Blocker AJ.** 2007. The type III secretion system needle tip complex mediates host cell sensing and translocon insertion. Mol Microbiol. **63:**1719-1730.

140. **Sani M, Allaoui A, Fusetti F, Oostergetel GT, Keegstra W, Boekema EJ.** 2007. Structural organization of the needle complex of the type III secretion apparatus of *Shigella flexneri*. Micron. **38:**291-301.

141. **Sani M, Botteaux A, Parsot C, Sansonetti P, Boekema EJ, Allaoui A.** 2007. IpaD is localized at the tip of the *Shigella flexneri* type III secretion apparatus. Biochim Biophys Acta. **1770:**307-311.

142. **Olive AJ, Kenjale R, Espina M, Moore DS, Picking WL, Picking WD.** 2007. Bile salts stimulate recruitment of IpaB to the *Shigella flexneri* surface, where it colocalizes with IpaD at the tip of the type III secretion needle. Infect. Immun. **75:**2626-2629.

143. **Espina M, Olive AJ, Kenjale R, Moore DS, Ausar SF, Kaminski RW, Oaks EV, Middaugh CR, Picking WD, Picking WL.** 2006. IpaD localizes to the tip of the type III secretion system needle of *Shigella flexneri*. Infect. Immun. **74:**4391-4400.

144. **Kolin A, Balasubramaniam V, Skredenske JM, Wickstrum JR, Egan SM.** 2008. Differences in the mechanism of the allosteric L-rhamnose responses of the AraC/XylS family transcription activators RhaS and RhaR. Mol Microbiol. **68:**448-461.

145. **Yang J, Hart E, Tauschek M, Price GD, Hartland EL, Strugnell RA, Robins-Browne RM.** 2008. Bicarbonate-mediated transcriptional activation of divergent operons by the virulence regulatory protein, RegA, from *Citrobacter rodentium*. Mol Microbiol. **68:**314-327.

146. **Abuaita BH, Withey JH.** 2009. Bicarbonate Induces *Vibrio cholerae* virulence gene expression by enhancing ToxT activity. Infect. Immun. **77:**4111-4120.

147. **Koppolu V, Osaka I, Skredenske JM, Kettle B, Hefty PS, Li J, Egan SM.** 2013. Small-Molecule Inhibitor of the *Shigella flexneri* Master Virulence Regulator VirF. Infect. Immun. **81:**4220-4231.

148. **Skredenske JM, Koppolu V, Kolin A, Deng J, Kettle B, Taylor B, Egan SM.** 2013. Identification of a small-molecule inhibitor of bacterial AraC family activators. J Biomol Screen. **18:**588-598.

149. **Neidhardt FC, Bloch PL, Smith DF.** 1974. Culture medium for enterobacteria. J Bacteriol. **119:**736-747.

150. **Stewart GS, Lubinsky-Mink S, Jackson CG, Cassel A, Kuhn J.** 1986. pHG165: a pBR322 copy number derivative of pUC8 for cloning and expression. Plasmid. **15:**172-181.

151. **Backman K, Chen YM, Magasanik B.** 1981. Physical and genetic characterization of the *glnA--glnG* region of the *Escherichia coli* chromosome. Proceedings of the National Academy of Sciences of the United States of America. **78:**3743-3747.

152. **Peters JE, Thate TE, Craig NL.** 2003. Definition of the *Escherichia coli* MC4100 genome by use of a DNA array. J Bacteriol. **185:**2017-2021.

153. **Miller JH.** 1992. A Short Course in Bacterial Genetics: A Laboratory Manual for *Escherichia coli* and Related Bacteria. . Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

154. **LaPointe CF, Taylor RK.** 2000. The type 4 prepilin peptidases comprise a novel family of aspartic acid proteases. J Biol Chem. **275:**1502-1510.

155. **Bhende PM, Egan SM.** 1999. Amino acid-DNA contacts by RhaS: an AraC family transcription activator. J Bacteriol. **181:**5185-5192.

156. **Miller JH.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

157. **Koppolu V, Osaka I, Skredenske JM, Kettle B, Hefty PS, Li J, Egan SM.** 2013. Small Molecule Inhibitor of Shigella flexneri Master Virulence Regulator, VirF. Infect. Immun.

158. **Jullien N, Herman JP.** 2011. LUEGO: a cost and time saving gel shift procedure. Biotechniques. **51:**267-269.

159. **Roy A, Kucukural A, Zhang Y.** 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. **5:**725-738.

160. **Zhang Y.** 2008. I-TASSER server for protein 3D structure prediction. BMC bioinformatics. **9:**40.

161. **Childers BM, Weber GG, Prouty MG, Castaneda MM, Peng F, Klose KE.** 2007. Identification of residues critical for the function of the *Vibrio cholerae* virulence regulator ToxT by scanning alanine mutagenesis. J Mol Biol. **367:**1413-1430.

162. **Kwon HJ, Bennik MH, Demple B, Ellenberger T.** 2000. Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA. Nat Struct Biol. **7:**424-430.

163. **Gillette WK, Martin RG, Rosner JL.** 2000. Probing the *Escherichia coli* transcriptional activator MarA using alanine-scanning mutagenesis: residues important for DNA binding and activation. J Mol Biol. **299:**1245-1255.

164. **Tobes R, Ramos JL.** 2002. AraC-XylS database: a family of positive transcriptional regulators in bacteria. Nucleic Acids Res. **30:**318-321.

165. **Luscombe NM, Laskowski RA, Thornton JM.** 2001. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res. **29:**2860-2874.

166. **Eustance RJ, Bustos SA, Schleif RF.** 1994. Reaching out. Locating and lengthening the interdomain linker in AraC protein. J Mol Biol. **242:**330-338.

167. **Singh J, Thornton JM.** 1992. Atlas of protein side-chain interactions. IRL Press at Oxford University Press, Oxford ; New York.

168. **Gaston K, Bell A, Kolb A, Buc H, Busby S.** 1990. Stringent spacing requirements for transcription activation by CRP. Cell. **62:**733-743.

169. **Porter ME, Smith SG, Dorman CJ.** 1998. Two highly related regulatory proteins, *Shigella flexneri* VirF and enterotoxigenic *Escherichia coli* Rns, have common and distinct regulatory properties. FEMS Microbiol Lett. **162:**303-309.

170. **Dorman CJ.** 1992. The VirF protein from *Shigella flexneri* is a member of the AraC transcription factor superfamily and is highly homologous to Rns, a positive regulator of virulence genes in enterotoxigenic *Escherichia coli*. Mol Microbiol. **6:**1575.

171. **Hale TL.** 1991. Genetic basis of virulence in *Shigella* species. Microbiol Rev. **55:**206-224.

172. **Porter ME, Dorman CJ.** 2002. In vivo DNA-binding and oligomerization properties of the *Shigella flexneri* AraC-like transcriptional regulator VirF as identified by random and site-specific mutagenesis. J Bacteriol. **184:**531-539.

173. **Dorman CJ.** 2004. H-NS: a universal regulator for a dynamic genome. Nat Rev Microbiol. **2:**391-400.

174. **Savelkoul PH, Willshaw GA, McConnell MM, Smith HR, Hamers AM, van der Zeijst BA, Gaastra W.** 1990. Expression of CFA/I fimbriae is positively regulated. Microb Pathog. **8:**91-99.

175. **Gaastra W, Hamers, A.M., Jordi, B.J.A.M., Savelkoul, P.H.M., Willshaw, G.A., McConnell, M.M., Kusters, J.G., van Vliet, A.H.M. and van der Zeijst, B.A.M.** 1991. Regulation of expression of fimbriae of human enterotoxigenic *Escherichia coli*, p.

61-70. *In* Wädstrom T (ed.), Molecular Pathogenesis of Gastrointestinal Infections. Plenum Press, New York.

176. **Munson GP, Holcomb LG, Scott JR.** 2001. Novel group of virulence activators within the AraC family that are not restricted to upstream binding sites. Infect. Immun. **69:**186-193.

177. **Gorden J, Small PL.** 1993. Acid resistance in enteric bacteria. Infect. Immun. **61:**364-367.

178. **Labrec EH, Schneider H, Magnani TJ, Formal SB.** 1964. Epithelial Cell Penetration as an Essential Step in the Pathogenesis of Bacillary Dysentery. J Bacteriol. **88:**1503-1518.

179. **Evans DG, Satterwhite TK, Evans DJ, Jr., DuPont HL.** 1978. Differences in serological responses and excretion patterns of volunteers challenged with enterotoxigenic *Escherichia coli* with and without the colonization factor antigen. Infect. Immun. **19:**883-888.

180. **Guth BE.** 2000. Enterotoxigenic *Escherichia coli*--an overview. Memorias do Instituto Oswaldo Cruz. **95 Suppl 1:**95-97.

181. **Pope LM, Reed KE, Payne SM.** 1995. Increased protein secretion and adherence to HeLa cells by Shigella spp. following growth in the presence of bile salts. Infect. Immun. **63:**3642-3648.

182. **Lan R, Reeves PR.** 2002. *Escherichia coli* in disguise: molecular origins of Shigella. Microbes Infect. **4:**1125-1132.

183. **Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM.** 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. **57:**81-91.

184. **Lukjancenko O, Wassenaar TM, Ussery DW.** 2010. Comparison of 61 sequenced *Escherichia coli* genomes. Microb Ecol. **60:**708-720.

185. **Gupta S, Chowdhury R.** 1997. Bile affects production of virulence factors and motility of *Vibrio cholerae*. Infect. Immun. **65:**1131-1134.

186. **Schuhmacher DA, Klose KE.** 1999. Environmental signals modulate ToxT-dependent virulence factor expression in *Vibrio cholerae*. J Bacteriol. **181:**1508-1514.

187. **Sansonetti PJ.** 2004. War and peace at mucosal surfaces. Nat Rev Immunol. **4:**953-964.

188. **Forstner G, Wesley A, Forstner J.** 1982. Clinical aspects of gastrointestinal mucus. Adv Exp Med Biol. **144:**199-224.

189. **Allen A, Bell A, Mantle M, Pearson JP.** 1982. The structure and physiology of gastrointestinal mucus. Adv Exp Med Biol. **144:**115-133.

190. **Allen A.** 1978. Structure of gastrointestinal mucus glycoproteins and the viscous and gel-forming properties of mucus. Br Med Bull. **34:**28-33.

191. **Podolsky DK, Isselbacher KJ.** 1983. Composition of human colonic mucin. Selective alteration in inflammatory bowel disease. J Clin Invest. **72:**142-153.

192. **Vercellotti JR, Salyers AA, Wilkins TD.** 1978. Complex carbohydrate breakdown in the human colon. Am J Clin Nutr. **31:**S86-S89.

193. **Peekhaus N, Conway T.** 1998. What's for dinner?: Entner-Doudoroff metabolism in Escherichia coli. J Bacteriol. **180:**3495-3502.

194. **Montagne L, Toullec R, Lalles JP.** 2000. Calf intestinal mucin: isolation, partial characterization, and measurement in ileal digesta with an enzyme-linked immunosorbent assay. J Dairy Sci. **83:**507-517.

195. **Chang DE, Smalley DJ, Tucker DL, Leatham MP, Norris WE, Stevenson SJ, Anderson AB, Grissom JE, Laux DC, Cohen PS, Conway T.** 2004. Carbon nutrition of *Escherichia coli* in the mouse intestine. Proceedings of the National Academy of Sciences of the United States of America. **101:**7427-7432.

196. **Fox JT, Drouillard JS, Shi X, Nagaraja TG.** 2009. Effects of mucin and its carbohydrate constituents on *Escherichia coli* O157 growth in batch culture fermentations with ruminal or fecal microbial inoculum. J Anim Sci. **87:**1304-1313.

197. **Englesberg E, Anderson RL, Weinberg R, Lee N, Hoffee P, Huttenhauer G, Boyer H.** 1962. L-Arabinose-sensitive, L-ribulose 5-phosphate 4-epimerase-deficient mutants of *Escherichia coli*. J Bacteriol. **84:**137-146.

198. **Yudkin MD.** 1969. Catabolite repression of the *lac* operon. Separate repression of two enzymes. The Biochemical journal. **114:**313-319.

199. **Lopez JM, Thoms B.** 1977. Role of sugar uptake and metabolic intermediates on catabolite repression in *Bacillus subtilis*. J Bacteriol. **129:**217-224.

200. **Jacob F, Monod J.** 1961. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol. **3:**318-356.

201. **Koehler TM, Dai Z, Kaufman-Yarbray M.** 1994. Regulation of the *Bacillus anthracis* protective antigen gene: CO2 and a trans-acting element activate transcription from one of two promoters. J Bacteriol. **176:**586-595.

202. **Ross RA, Onderdonk AB.** 2000. Production of toxic shock syndrome toxin 1 by *Staphylococcus aureus* requires both oxygen and carbon dioxide. Infect. Immun. **68:**5205-5209.

203. **Abe H, Tatsuno I, Tobe T, Okutani A, Sasakawa C.** 2002. Bicarbonate ion stimulates the expression of locus of enterocyte effacement-encoded genes in enterohemorrhagic *Escherichia coli* O157:H7. Infect. Immun. **70:**3500-3509.

204. **Hogan DL, Ainsworth MA, Isenberg JI.** 1994. Review article: gastroduodenal bicarbonate secretion. Aliment Pharmacol Ther. **8:**475-488.

205. **Fordtran JS, Locklear TW.** 1966. Ionic constituents and osmolality of gastric and small-intestinal fluids after eating. Am J Dig Dis. **11:**503-521.

206. **Ruppin H, Domschke, W., Soergel, K.H.** 1981. Diarrhea in disorders of intestinal transport. Georg Thieme Verlag Stuttgart, Germany.

207. **Pilonieta MC, Bodero MD, Munson GP.** 2007. CfaD-dependent expression of a novel extracytoplasmic protein from enterotoxigenic *Escherichia coli*. J Bacteriol. **189:**5060-5067.

208. **Jost BH, Adler B.** 1993. 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. Microb Pathog. **14:**481-488.

209. **Lee N, Francklyn C, Hamilton EP.** 1987. Arabinose-induced binding of AraC protein to *araI2* activates the *araBAD* operon promoter. Proceedings of the National Academy of Sciences of the United States of America. **84:**8814-8818.

210. **Carra JH, Schleif RF.** 1993. Variation of half-site organization and DNA looping by AraC protein. EMBO J. **12:**35-44.

211. **Bodero MD, Pilonieta MC, Munson GP.** 2007. Repression of the inner membrane lipoprotein NlpA by Rns in enterotoxigenic *Escherichia coli*. J Bacteriol. **189:**1627-1632.

212. **Griffith KL, Wolf RE, Jr.** 2002. A comprehensive alanine scanning mutagenesis of the *Escherichia coli* transcriptional activator SoxS: identifying amino acids important for DNA binding and transcription activation. J Mol Biol. **322:**237-257.

213. **Freeman AM, Mole BM, Silversmith RE, Bourret RB.** 2011. Action at a distance: amino acid substitutions that affect binding of the phosphorylated CheY response regulator and catalysis of dephosphorylation can be far from the CheZ phosphatase active site. J Bacteriol. **193:**4709-4718.

214. **Schleif R, Hess W, Finkelstein S, Ellis D.** 1973. Induction kinetics of the L-arabinose operon of *Escherichia coli*. J Bacteriol. **115:**9-14.

215. **Scherbakov DV, Garber, M. B.** 2000. Overlapping genes in bacterial and phage genomes. Mol Biol. **34:**485-495.

216. **Nakagawa S, Niimura Y, Miura K, Gojobori T.** 2010. Dynamic evolution of translation initiation mechanisms in prokaryotes. Proceedings of the National Academy of Sciences of the United States of America. **107:**6382-6387.

217. **Adhin MR, van Duin J.** 1990. Scanning model for translational reinitiation in eubacteria. J Mol Biol. **213:**811-818.

218. **Lesage P, Chiaruttini C, Graffe M, Dondon J, Milet M, Springer M.** 1992. Messenger RNA secondary structure and translational coupling in the Escherichia coli operon encoding translation initiation factor IF3 and the ribosomal proteins, L35 and L20. J Mol Biol. **228:**366-386.

219. **Chiaruttini C, Milet M, de Smit M, Springer M.** 1996. Translational coupling in the Escherichia coli operon encoding translation initiation factor IF3 and ribosomal proteins L20 and L35. Biochimie. **78:**555-567.

220. **Petersen C.** 1989. Long-range translational coupling in the rplJL-rpoBC operon of Escherichia coli. J Mol Biol. **206:**323-332.

221. **Childers BM, Cao X, Weber GG, Demeler B, Hart PJ, Klose KE.** 2011. N-terminal residues of the *Vibrio cholerae* virulence regulatory protein ToxT involved in dimerization and modulation by fatty acids. J Biol Chem. **286:**28644-28655.

222. **Wattiau P, Cornelis GR.** 1994. Identification of DNA sequences recognized by VirF, the transcriptional activator of the Yersinia yop regulon. J Bacteriol. **176:**3878-3884.

223. **Lambert de Rouvroit C, Sluiters C, Cornelis GR.** 1992. Role of the transcriptional activator, VirF, and temperature in the expression of the pYV plasmid genes of *Yersinia enterocolitica*. Mol Microbiol. **6:**395-409.

224. **Shakhnovich EA, Hung DT, Pierson E, Lee K, Mekalanos JJ.** 2007. Virstatin inhibits dimerization of the transcriptional activator ToxT. Proceedings of the National Academy of Sciences of the United States of America. **104:**2372-2377.

225. **Parra MC, Collins CM.** 2012. Mutational analysis of the N-terminal domain of UreR, the positive transcriptional regulator of urease gene expression. Microbiol Res. **167:**433-444.

226. **Poore CA, Coker C, Dattelbaum JD, Mobley HL.** 2001. Identification of the domains of UreR, an AraC-like transcriptional regulator of the urease gene cluster in *Proteus mirabilis*. J Bacteriol. **183:**4526-4535.

227. **Ruiz R, Marques S, Ramos JL.** 2003. Leucines 193 and 194 at the N-terminal domain of the XylS protein, the positive transcriptional regulator of the TOL meta-cleavage pathway, are involved in dimerization. J Bacteriol. **185:**3036-3041.

228. **Bustos SA, Schleif RF.** 1993. Functional domains of the AraC protein. Proceedings of the National Academy of Sciences of the United States of America. **90:**5638-5642.

229. **LaRonde-LeBlanc N, Wolberger C.** 2000. Characterization of the oligomeric states of wild type and mutant AraC. Biochemistry. **39:**11593-11601.

230. **Coburn PS, Baghdayan AS, Dolan GT, Shankar N.** 2008. An AraC-type transcriptional regulator encoded on the *Enterococcus faecalis* pathogenicity island contributes to pathogenesis and intracellular macrophage survival. Infect. Immun. **76:**5668-5676.

231. **Frota CC, Papavinasasundaram KG, Davis EO, Colston MJ.** 2004. The AraC family transcriptional regulator Rv1931c plays a role in the virulence of *Mycobacterium tuberculosis*. Infect. Immun. **72:**5483-5486.

232. **Hauser AR, Kang PJ, Engel JN.** 1998. PepA, a secreted protein of *Pseudomonas aeruginosa*, is necessary for cytotoxicity and virulence. Mol Microbiol. **27:**807-818.

233. **Bowser TE, Bartlett VJ, Grier MC, Verma AK, Warchol T, Levy SB, Alekshun MN.** 2007. Novel anti-infection agents: small-molecule inhibitors of bacterial transcription factors. Bioorg Med Chem Lett. **17:**5652-5655.

234. **Garrity-Ryan LK, Kim OK, Balada-Llasat JM, Bartlett VJ, Verma AK, Fisher ML, Castillo C, Songsungthong W, Tanaka SK, Levy SB, Mecsas J, Alekshun MN.** 2010. Small molecule inhibitors of LcrF, a *Yersinia pseudotuberculosis* transcription factor, attenuate virulence and limit infection in a murine pneumonia model. Infect. Immun. **78:**4683-4690.

235. **Grier MC, Garrity-Ryan LK, Bartlett VJ, Klausner KA, Donovan PJ, Dudley C, Alekshun MN, Tanaka SK, Draper MP, Levy SB, Kim OK.** 2010. N-Hydroxybenzimidazole inhibitors of ExsA MAR transcription factor in *Pseudomonas aeruginosa*: In vitro anti-virulence activity and metabolic stability. Bioorg Med Chem Lett. **20:**3380-3383.

236. **Hung DT, Shakhnovich EA, Pierson E, Mekalanos JJ.** 2005. Small-molecule inhibitor of *Vibrio cholerae* virulence and intestinal colonization. Science. **310:**670-674.

237. **Hurt JK, McQuade TJ, Emanuele A, Larsen MJ, Garcia GA.** 2010. High-throughput screening of the virulence regulator VirF: a novel antibacterial target for shigellosis. J Biomol Screen. **15:**379-387.

238. **Kim OK, Garrity-Ryan LK, Bartlett VJ, Grier MC, Verma AK, Medjanis G, Donatelli JE, Macone AB, Tanaka SK, Levy SB, Alekshun MN.** 2009. N-hydroxybenzimidazole inhibitors of the transcription factor LcrF in Yersinia: novel antivirulence agents. J Med Chem. **52:**5626-5634.