Molecular Mechanisms and Inhibition of Transcription Activation by Bacterial AraC Family Activator Proteins

By

Veerendra Koppolu

Submitted to the graduate degree program in the Department of 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 Egan

_____
Dr. P. Scott Hefty

_____
Dr. Liskin Swint-Kruse

_____
Dr. Mizuki Azuma

_____
Dr. Liang Tang

Date Defended: 10/25/2013

The Dissertation Committee for Veerendra Koppolu

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

Molecular Mechanisms and Inhibition of Transcription Activation by Bacterial AraC Family Activator Proteins

_____

Chairperson: Dr. Susan Egan

Date approved: 11/04/2013

Dedication

This work is dedicated to my father, Brahmaiah Koppolu

ABSTRACT

AraC family proteins are transcriptional regulators that are defined by the presence of a conserved DNA binding domain (DBD). My research focused on three AraC family activators: Rns (activator of virulence genes in diarrhea-causing ETEC), VirF (activator of virulence genes in diarrhea-causing *Shigella*) and RhaR (activator of L-rhamnose catabolic operons in *Escherichia coli*). With the ultimate goal of discovery of novel antibacterial agents that inhibit the AraC family proteins, here I have investigated the molecular mechanism of transcription activation by Rns and RhaR. Site-directed mutagenesis of residues in the ETEC Rns N-terminal domain (NTD) identified three residues (N15, N16 and I17) that are required for the transcription activation function of Rns. Site-directed mutagenesis of residues in the Rns DBD (predicted to be contacted by the NTD residues) identified three residues (K216, Y251 and G252) that are required for transcription activation, and one residue (H250) that is required for both DNA binding and transcription activation. We propose that transcription activation by Rns involves contacts between RS2 and AS2 region residues and these contacts may impart the structure or dynamics required by Rns to activate transcription.

In RhaR, I investigated the role of the RhaR Arm in transmission of the signal that effector (L-rhamnose) is bound from the NTD to the DBD, converting RhaR to its activating state. Site-directed mutagenesis results suggested that the RhaR Arm is involved in maintaining RhaR in its non-activating state. Our results suggest that residue L35 in the Arm makes inter-domain interactions with the RhaR DBD to reduce transcription activation by RhaR in the absence of L-rhamnose.

To identify novel agents that target AraC family proteins, I tested the small molecule SE-1, which our lab identified as an effective inhibitor of the AraC family proteins RhaS and RhaR. Despite limited sequence identity, SE-1 was also shown to inhibit VirF and Rns activity in cell-based assays in *E. coli*. I showed that SE-1 blocked *in vitro* DNA binding by VirF and Rns, and expression of VirF-dependent virulence genes in *Shigella*. A collaborator showed that SE-1 inhibited invasion of *Shigella* into eukaryotic host cells. SE-1 did not detectably inhibit the growth or metabolism of the bacterial or eukaryotic host cells, respectively, indicating that the inhibition of invasion was not due to general toxicity. Overall, SE-1 appears to exhibit selectivity toward AraC family proteins, and has potential to be developed into a novel antibacterial agent.

Acknowledgements

I take this opportunity to acknowledge my mentor Dr. Susan Egan for her continuous guidance throughout the five years of my research in her laboratory. All the success I achieved in this laboratory will be truly credited to her. It would have been impossible for me to finish my projects and dissertation without her outstanding supervision. She spent enormous number of hours in training me on research techniques, problem solving and most importantly in improving my presentation skills. I feel short of words to express my gratitude towards her.

With sincere gratitude I acknowledge my indebtedness to my committee members Dr. Scott Hefty and Dr. Liskin-Swint-Kruse for their comments, suggestions and guidance over the years. I also thank my other committee members Dr. Liang Tang and Dr. Mizuki Azuma for extending their help when required. I record my sincere thanks to my colleagues and friends Jeff, James, Gurpreet, Jiaquin, Bria, Ichie, Iera, Sunil, Sunder, Frances and Maged for helping me either directly or indirectly during my PhD. In particular, I am thankful to Jeff and James for their unconditional help and willingness to answer my questions all the time. I am also thankful to John Connolly for help in departmental matters. Lawrence gave me good memories over the past few years. I made a huge list of friends here and I thank all of them for giving me a wonderful company over the years.

All my achievements would be meaningless if I fail to acknowledge my parents, grandparents, brother and sister-in-law. Your love, affection and constant support has made me what I am today. I am sure you will be proud of my achievements. I can't really emphasize how big a role my wife, Veneela, had in finishing this dissertation. She clearly understood the challenges one has to face during their graduate career and wholeheartedly supported me with a lot of patience. Thanks for thinking my success as your success and I feel incredibly lucky to have you in my life.

Table of Contents

Chapter 1

INTRODUCTION

Significance. Approval of the first antibiotic Penicillin in 1940s and subsequent release of several classes of antibiotics in 1950s and 1960s have dramatically reduced bacterial infections worldwide (1, 2). However, indiscriminate use of antibiotics led to rapid development of resistance by many pathogenic bacteria (3). This unfortunate situation has been exacerbated by the decrease in antibiotic research, which was evident from the fact that no new class of antibiotic was released between 1962 and 2000. In recent years, the development of new antibiotics has generally been limited to modifications of existing antibiotics to improve their efficacy (2, 4-6). A recent report showed that only six antibiotics have been approved for public use in USA since 2003 (7)

Until recently, the targets for antibiotics are essential bacterial proteins that have roles in important cellular processes like cell wall synthesis, DNA replication and protein synthesis (8-15). Targeting of these essential cellular processes by antibiotics can either kill or inhibit the bacterial growth, and thus put pressure on the bacteria to select for resistant population (9). In order to successfully combat the antimicrobial resistance we need new strategies to develop antibiotics. One such approach has been suggested recently and involves targeting of bacterial virulence factors that are required to cause successful bacterial infection (9, 10). Given that targeting virulence proteins does not kill the bacteria directly this strategy might decrease the rate at which bacteria develop resistance. In addition, inhibition of virulence factors could allow the immune system to act on bacterium to clear the infection (9, 16-18). As an alternative to targeting of virulence proteins, strategies that target the AraC family virulence activators were thought to have potential to develop into antibacterial agents (19-23). In many pathogenic bacteria it has been shown that these AraC family proteins are required for the bacteria to cause infection (24-28). Further, several recent studies that targeted AraC family proteins showed a

reduction in the bacterial virulence and thus suggesting that AraC family proteins can be developed as antibacterial targets (19-23). In addition, AraC family virulence activators in many pathogenic bacteria show sequence similarities, and therefore compounds that target AraC family proteins could be used as broad-spectrum antibiotics to treat multiple infections (29-34).

General characteristics of AraC family proteins. AraC family proteins constitute one of the largest family of bacterial regulatory proteins with more than 1900 proteins identified to be in this family so far (35). Proteins in the AraC family are present in ~70% of the sequenced bacterial genomes and mainly regulate transcription (35). Protein members of this family are defined by a sequence similarity within 99 amino acids that constitutes the DNA binding domain (36). Multiple AraC family proteins are present in each bacterium, with *E. coli* K12 having 27 AraC family proteins identified so far (37). Although majority of the proteins in this family are still uncharacterized, proteins characterized so far are found to have roles in control of expression of genes involved in various biological processes including carbon metabolism, stress responses, and most importantly virulence (37). Majority of the well-studied AraC family proteins are transcriptional activators, with the exception of CelD (transcriptional repressor) (38), AraC, YbtA and Rns (both activators and repressors depending on the promoter regions where they bind) (39-44).

Depending on the biological processes AraC family proteins regulate, these proteins can be categorized into three different classes. AraC family proteins that are found to be involved in regulation of virulence genes include Rns/CfaD (pili forming genes in Enterotoxigenic *E. coli*) (41), VirF (genes involved in invasion and cell-to-cell spread of *Shigella* spp. in human intestine) (45-49), PerA (bundle-forming pili genes in Enteropathogenic *E. coli*) (50), RegA (attachment and effacement of *Citrobacter rodentium* to mice intestinal epithelium) (51), MxiE [expression

of Type3 secretion system (T3SS) effectors after invasion of *Shigella* spp. into host intestinal epithelial cells] (52), ToxT (Cholera toxin and toxin-coregulated pilus in *Vibrio cholerae*) (53), ExsA (exoenzyme S in *Pseudomonas aeruginosa*) (54), UreR (Urease in various bacterial spp.) (55), InvF (T3SS genes in *Salmonella typhimurium*) (56). Some of these regulatory proteins respond to specific effector molecules (RegA - bicarbonate ions; ToxT - fatty acids and bicarbonate ions; UreR – urea) (57-60) while some of these respond to physical stimuli like temperature, pH, osmolarity of the medium and Ca^{2+} concentration (36).

AraC family proteins that are involved in carbon catabolism include RhaS and RhaR (sugar L-rhamnose in *E. coli* K12) (61), AraC (L-arabinose in *E. coli* K12) (62), MmsR (amino acid valine in *Pseudomonas aeruginosa*) (63), XylS (degradation of toluene in *Pseudomonas putida*) (64), MelR (melobiose in *E. coli* K12) (65) and many more . Members of this group are found to activate transcription in response to binding of effector molecules. AraC family proteins that regulate stress response genes include Rob (antibiotic resistance and heavy metal resistance in *E. coli* K12) (66), SoxS (superoxide response regulon in *E. coli* K12) (67), MarA (antibiotic resistance in *E. coli* K12) (68), Ada (alkylating agents in *E. coli* and *Salmonella typhimurium* and *Mycobacterium tuberculosis*) (69, 70) and many more.

Structural domains and biochemical functions of AraC family proteins. Members of the AraC family proteins are generally comprised of two structural domains: a less conserved N-terminal domain (NTD) and a conserved DNA binding domain (DBD). However, given the large size of the family there are few members (~5%) that do not fit into this classification and found to have a single domain (36, 71). The two domains in the AraC family proteins can be connected by a flexible linker (72). In AraC, recently it was found that substitutions in linker affect the affinity of AraC to bind to its half-sites on DNA. Further, their results suggest that AraC linker is

more likely involved in interactions between NTD and DBD of AraC (73). In RhaS and RhaR, single alanine substitution in the linker region have at most smaller effects on L-rhamnose response, suggesting that linker may not play a crucial role in RhaS and RhaR activation (74).

N-terminal domain (NTD). NTD is a less conserved domain among AraC family proteins, although NTDs of some subgroups with in the family are conserved (36, 71, 75). For examples, NTDs of some protein such as AraC, RhaS and RhaR share sequence similarity to each other over their entire length. Unlike the DBD, N-terminal domain of the AraC family proteins is not studied extensively. Studies on AraC, ToxT, RhaS and RhaR showed that NTD is involved in dimerization and binding of effector molecules (76-78). Conclusive evidence for the dimerization and effector binding by AraC is obtained from the crystal structure of this protein solved in the presence of its effector L-arabinose. Evidence for the dimerization of *Vibrio cholerae* virulence activator ToxT came from dominant inhibition, LexA fusion, and two-hybrid analysis (20, 78) while evidence for effector binding was found from the crystal structure of protein solved in the presence of fatty acid effector (79). Evidence for the dimerization of RhaS and RhaR came from dominant negative experiments done by former graduate student Ana Kolin from my lab (76) (Ana Kolin, G.K. Hanjun and Susan Egan unpublished result). In dominant negative experiments, first she overexpressed RhaS-NTD and RhaR-NTD in strains that encode wild-type RhaS and RhaR leading to a dominant negative phenotype (a phenotype of decreased RhaS and RhaR activity because of formation of heterodimer between NTD and full-length protein). Then she overexpressed different RhaS-NTD and RhaR-NTD variants that have mutations in the predicted dimerization residues and tested for the loss-of- dominant negative phenotype. This screen resulted in identifications of residues in the NTDs of RhaS and RhaR that might have role in dimerization.

Conserved DNA binding domain (DBD). As mentioned before AraC family proteins are characterized by the presence of DBD which is ~99 amino acids long and share sequence similarity within the family. The DBD is generally present at the C-terminus of the protein (although Rob and CfaR from *E. coli* have DBD at the N-terminus of the protein). DBD was found to have a role in DNA binding and transcription activation through contacts with RNA polymerase (RNAP) (35, 36, 71, 76, 77, 80-88). For members that have only the DBD (MarA and SoxS), it was shown that this domain is sufficient for transcription activation (89, 90). Secondary structure predictions and availability of full-length structures of MarA, Rob, (both were solved as co-crystals with DNA), ToxT (solved in effector bound state) and DBD of AraC have shown that DBD has two independent helix-turn-helix motifs (HTH1 and HTH2) connected by an α-helix. Sequence comparison of both the HTH motifs indicated that HTH2 is more conserved, with certain residues showing more than 90% identity (36). Genetic and biochemical analysis showed that many members of this family (MarA, AraC, Rns, PerA, RhaS, XylS and VirF) use both the HTH motifs to make base specific contacts with DNA (91-97). Structure of MarA specifically showed that a recognition helix from each helix-turn-helix motif is placed in the major groove of DNA (91). Genetic and biochemical analysis have revealed that DBDs have a role in interacting with σ^{70} of RNA polymerase (RNAP) to activate transcription, especially the transcription activators that bind at site on DNA overlapping -35 hexamer (80-83).

Mechanism of transcription activation by AraC and ToxT

AraC. AraC is the founding member of the AraC family proteins and is by far one of the very well characterized protein in the family. AraC is a dimer and each monomer of AraC binds to a single molecule of sugar L-arabinose (77, 98). AraC activates the transcription of genes in the L-arabinose operon in *E. coli* in the presence of its effector L-arabinose (99). In AraC, an Arm

region (comprising of residues 7-14) was found to be required to keep AraC in a non-activation state in the absence of L-arabinose (100). Genetic and biochemical data showed that the Arm in the AraC-NTD makes differential contacts with DBD and arabinose in the absence and presence of arabinose, respectively (100, 101). The role of these structural changes in the Arm was explained by the proposed "light switch" signaling mechanism of transcription activation by AraC (79, 101-106) (Fig. 1). According to this mechanism, contacts between the N-terminal Arm and the DBD in the absence of L-arabinose constrain the DBD, such that the DBD of one monomer binds at a distantly spaced half-site (*araO*$_2$) from the DBD of another AraC monomer (bound at half-site at *araI*$_1$), and thus repress transcription of the *araBAD* operon (101, 103, 106-111) (Fig. 1). When arabinose is bound to the AraC-NTD, N-terminal Arm moves to contact arabinose and thus releases the constraint on the AraC-DBD (105, 106, 108). In this arabinose bound state, the two domains of the AraC are flexibly connected and adopt a conformation in which both the DBDs of both the AraC monomers bind to closely spaced half sites (*araI*$_1$ and *araI*$_2$) on the *araBAD* promoter and thus activate transcription (Fig. 1). However, very recently it was found that NTD makes inter-domain contacts with the DBD even in the presence of arabinose, and these contacts are different from the contacts made in the absence of arabinose (112).

ToxT. ToxT is an AraC family virulence activator of cholera toxin and toxin coregulated pilus encoding genes in *Vibrio cholerae* (53). ToxT behaves as either monomer or dimer depending on

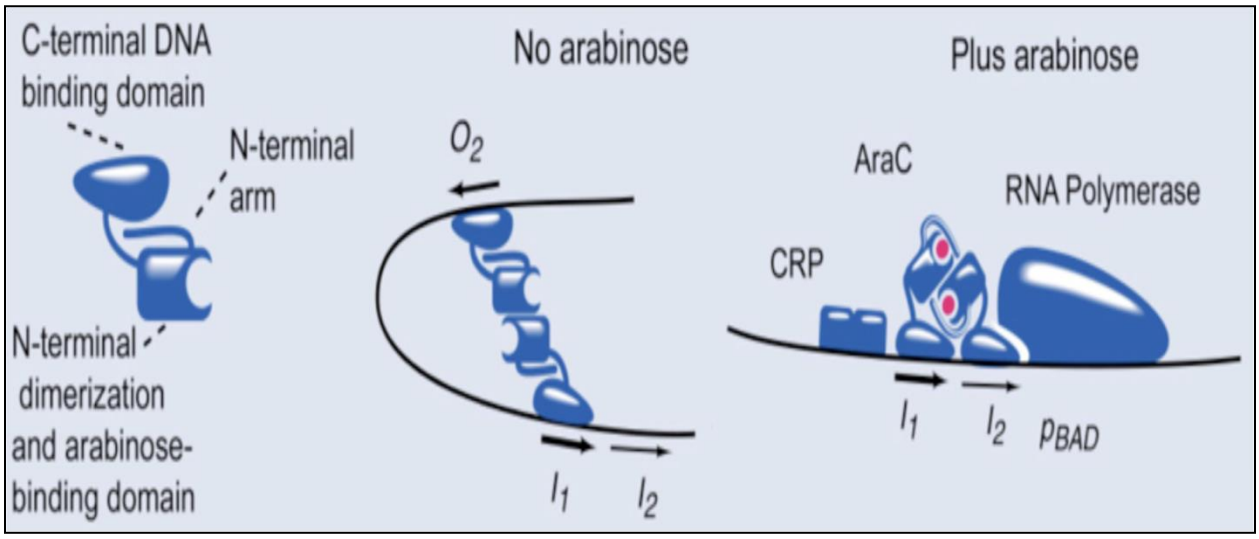


Fig. 1. AraC light-switch mechanism. In the (-) arabinose state, AraC N-terminal Arm in the NTD interacts with the DBD. In this conformation, the DBD binds at two distantly spaced half-sites (*araO*$_2$ and *araI*$_1$). In the (+) arabinose state, the N-terminal Arm bends over the arabinose and thus releases the DBD. DBDs of both the AraC monomers bind at adjascent half-sites (*araI*$_1$ and *araI*$_2$), which activates transcription from the arabinose operon promoter region $P_{BAD.}$ Figure is from the reference (113).

the promoter region where it binds. The crystal structure of the ToxT is available and is the only available structure among the virulence activators of the AraC family and serves as a model for the other AraC family virulence activators (Fig. 2). The crystal structure of ToxT was solved with fatty acid effector bound in an effector binding pocket in the NTD (79). ToxT crystal structure showed a very large (~2000 $Å^2$) polar inter-domain interface, with many polar and few hydrophobic interactions between NTD and DBD (79). ToxT responds to the bile salts and unsaturated fatty acid effectors such as arachidonic, linoleic, oleic, and plamitoleic acid (59, 79, 114). When bound by these effector molecules ToxT does not interact with DNA and therefore does not activate transcription (59, 79, 114).

To explain the mechanism of transcription activation by ToxT, a model has been proposed based on the structural and genetic analysis (79, 114). According to this model, ToxT is in "closed" conformation in the presence of effector and in "open" conformation in the absence of effector. In the presence of effector, NTD residues that surround the entrance to the ToxT effector-binding pocket, as well as the fatty acid effector itself, contact the DBD and keep the ToxT-DBD in an orientation that is not favorable to DNA binding (closed conformation) (79). These effector dependent inter-domain interactions between the ToxT NTD and DBD are a likely mechanism for transmission of the effector-binding status in ToxT. The model propose that in the absence of effector, the inter-domain interactions would be lost between NTD and DBD leading to destabilization of the closed complex into an open conformation that is capable of binding DNA and activating transcription. Alternatively, given the large ToxT inter-domain interface (79) and the presence of inter-domain contacts in AraC in both (+) and (-) arabinose state (112) we hypothesize that the ToxT inter-domain contacts may be altered/rearranged, but not eliminated, in its activation state.

AraC family virulence regulator Rns in Enterotoxigenic *E. coli*.

Enterotoxigenic *Escherichia coli* (ETEC) is a common cause of infantile and traveler's diarrhea (115, 116), and is most commonly acquired through contaminated food and water. It is estimated that ETEC causes nearly a billion cases of diarrheal disease and 300,000 – 500,000 deaths worldwide each year, mostly in children under the age of five (117). In the United States, ETEC infections are most often acquired during travel to endemic areas or in sporadic food-borne outbreaks (118-122). Worldwide, there are an estimated 40 million cases of traveler's diarrhea, of which approximately one-third are caused by ETEC (118).

Pathogenicity of ETEC strains is dependent on production of either one or both enterotoxins called heat-stable (ST) and heat-labile (ST) toxins (123, 124). These enterotoxins enter the host intestinal epithelial cells through receptor mediated endocytosis and causes opening of ion channels, which leads to loss of water and ions (Cl^-, K^+, Na^+, HCO_3^-) and thus causing profuse watery diarrheal symptoms (125-130). Effective release of the enterotoxins into the host intestinal epithelial cells occurs upon successful intestinal colonization by the bacteria and requires adherence of bacteria to the intestinal epithelial cells (131). Adherence of bacteria to the target epithelial cells is mediated through proteinaceous surface structures called pili/fimbrial colonization factors (CFs) (132, 133) and the expression of many of these CFs are activated by transcriptional activator Rns or its closest homolog CfaD (40-42).

At least 25 different fimbrial CFs have been identified in different ETEC strains, although most strains express only two or three CFs (134-138). The most commonly expressed CFs are CFA/I, CS1, CS2, CS3, CS4, CS5 and CS6 (133, 139-143), out of which regulation of CFA/I, CS1 and CS2 are well studied so far. ETEC strains lacking fimbriae are unable to

colonize intestine and cause disease (132), indicating the central role fimbriae has in the overall ETEC pathogenesis. In well-studied infectious ETEC strains, Rns is found to activate the expression of fimbrial types and thereby considered as a major virulence regulator (41, 42). The fimbrial types Rns activates include CS1, CS2, CS3, CS4, CS14, CS17 and CS19 (41, 42). An Rns homologous protein CfaD/CfaR is required for the expression of CFA/I fimbriae (40). CfaD/CfaR is ~95% identical (251 out of 265 residues) and is functionally interchangeable with Rns, and recognizes the same binding sites as Rns (40). Limited information is available regarding the regulation of expression of other fimbrial types and it is possible that Rns/CfaD may enhance the transcription of these fimbrial types as well.

In addition to the activation of fimbrial genes, Rns positively regulates the expression of its own gene (144) directly by binding at three sites in the promoter region of *rns*. The regions where Rns binds are centered at -227, +43 and +82 relative to transcription start site (TSS) (144). Rns activates the expression of several other putative virulence factor genes like *cexE* (encodes an uncharacterized extra-cytoplasmic protein) (145) and *yisS* (unknown function) (146). Rns represses the expression of a periplasmic protein NlpA by binding (at +5 to +40 relative to TSS) and preventing the formation of an open complex by RNAP at the *nlpA* promoter region (39).

DNaseI footprinting has revealed Rns binding site immediately upstream of -35 region relative to TSS at promoters of CfaA, CS1 (P*coo* promoter), CS17, CS19 and PCFO71(42, 147), suggesting that Rns may activate transcription by directly contacting RNAP. In addition to binding site near -35 region, a second binding site was identified at all the promoter regions of these fimbrial types (CfaA at region covering -88; CS1 at -144; for CS17 and PCFO71 at -109.5; CS19 at -108.5) (42, 147). It was found that each site has an additive effect on Rns dependent transcription and therefore is required for full activation of Rns. A highly unconventional

arrangement of binding sites is found to be involved in Rns autoactivation from its own promoter. The binding sites at *rns* promoter regions are centered at -227, +43 and +82 relative to TSS. Presence of downstream binding sites is commonly associated with transcription repressors. However, it was found that activation of *rns* requires the presence of upstream binding site (-227) and at least one of the two downstream binding sites (144).

Rns homologs have been identified in several ETEC strains and other enteric bacterial pathogens. Proteins with strong homology to Rns are CsvR and FapR from ETEC (33, 34), AggR of enteroaggregative *E. coli* (32), PerA of enteropathogenic *E. coli* (31), VirF of *Shigella* (30) and ToxT of *Vibrio cholerae* (29). Among these, CsvR, AggR, PerA and ToxT regulate the expression of fimbrial genes (29, 31-34). VirF of *Shigella* activates the expression of *icsA* and *virB* genes, that have roles in invasion and cell-to-cell spread of bacteria in host epithelial cells (45-49, 148). Activation by Rns and the above mentioned homologs are sensitive to temperature and activate the expression from promoters that are repressed by nucleoid associated global transcriptional regulator H-NS at low temperatures (149-153)

Mechanism of transcription activation by Rns. Rns activates the transcription of many genes encoding the fimbrial colonization factors (CFs) required for ETEC infection. However, the molecular mechanism of transcription activation by Rns is not clearly understood. Like AraC family proteins, the DBD of Rns has two predicted HTH motifs through which it binds to DNA. Uracil interference studies by Munson *et al* showed that Rns contacts the two major grooves of the DNA (147, 154) to activate transcription. Pentapeptide insertional mutagenesis experiments revealed that Rns uses both HTH motifs for making DNA contacts and thereby activating CS1 fimbrial gene expression (93). Like many DBDs of AraC family proteins, DBD of Rns is thought

to be involved in making RNAP contacts, although the region involved in making RNAP contacts has not been identified.

Unlike the DBD, the role of the NTD in Rns function is largely unknown. It was not fully known whether Rns acts primarily as a monomer or dimer as the previous reports were quite contradictory to each other (155, 156) Additionally, there is no evidence that Rns responds to exogenous effector molecules (156) (although structural prediction program I-TASSER (157) predicted the presence of a large effector binding pocket similar to effector binding pocket of ToxT structure). Nonetheless, both in vivo and in vitro experiments with N-terminal deletions of Rns indicate that the NTD is required for transcription activation (156). An Rns deletion construct with first 61 amino acids deleted has shown a complete loss of DNA binding and transcription activation or repression at their respective promoter regions (156). A region of residues I12-M18 in Rns-NTD are highly conserved among the closest homologs of Rns and shares a sequence identity of ~74% compared to the overall NTD identity of 26% among these proteins. This high percentage identity suggests an important role of this region in overall function of Rns. In further support of this hypothesis, Munson *et al* have isolated two random mutations at the N-terminus of the protein (I14T and N16D) whose activities were decreased dramatically at *rns* promoter indicating the importance of I14 and N16 residues in the transcription activation (156). They found that these mutations have disparate effects at other Rns regulated promoters and suggested that these residues may interact with DBD in a manner to activate transcription (156).

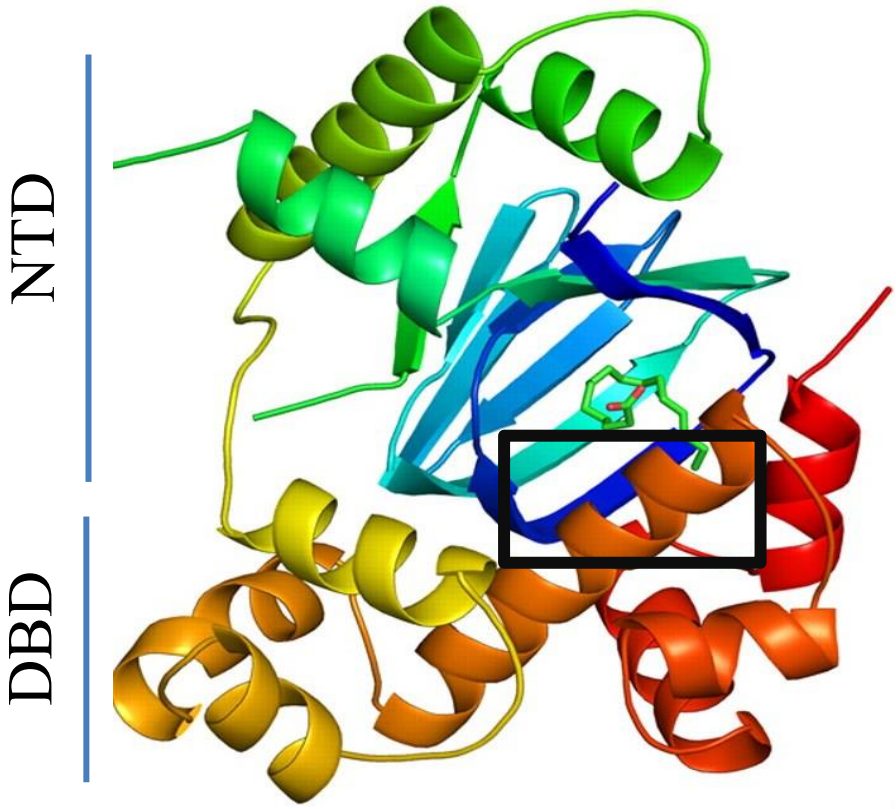


Fig. 2. ToxT structure (79)**.** Ribbon diagram of the ToxT with the effector palmitoleic acid (shown in stick form) bound in the NTD. The models are colored with rainbow effect with the blue at the N-terminus and red at the C-terminus. Rectangular black box represents the interface of the potential inter-domain interactions. The structure of ToxT serves as model structure for Rns.

AraC family virulence regulator VirF in Shigella spp

Shigella is a major cause of bacillary dysentery (shigellosis) in humans, a disease that is characterized by a short period of watery diarrhea with intestinal cramps, followed by bloody mucoid stools. Various pathogenic species of *Shigella*, such as *S. flexneri, S. dysenteriae, S. sonnei and S. boydii* cause shigellosis and are responsible for 165 million illnesses and over 1.1 million deaths worldwide each year, with 70% of the deaths occurring in children under the age of five (158, 159). Among *Shigella* species, *S. flexneri* causes more mortality than any other species, with as few as 100 cells being sufficient to cause disease (160). All these species carry a large 220 kb virulence plasmid that contains majority of the virulence genes required for the successful infection of the bacteria into colonic and rectal epithelial cells. These include genes required for the formation of the type 3 secretion system (T3SS) machinery, invasion of *Shigella* into host epithelial cells, and cell-to-cell spread (45-49).

Expression of majority of the plasmid encoded virulence genes are regulated at the transcriptional level by an AraC family transcriptional activator VirF. Expression of VirF has been shown to be temperature dependent, with three- to four-fold lower expression at 30^{o}C than at 37^{o}C (148, 161). At 37^{o}C, VirF (96) and activates the expression of a regulatory cascade of genes involving *icsA* and *virB* genes, which encode IcsA and VirB virulence proteins, respectively (Fig. 3). VirB in turn activates the expression of many virulence-associated genes (located in *mxi*, *spa* and *ipa* operons) (148) responsible for the formation of the T3SS machinery, invasion and cell-to-cell spread (45-49).

Transcription of VirF and its direct targets IcsA and VirB are under the negative control of nucleoid associated global regulator H-NS. H-NS negatively regulates these genes at

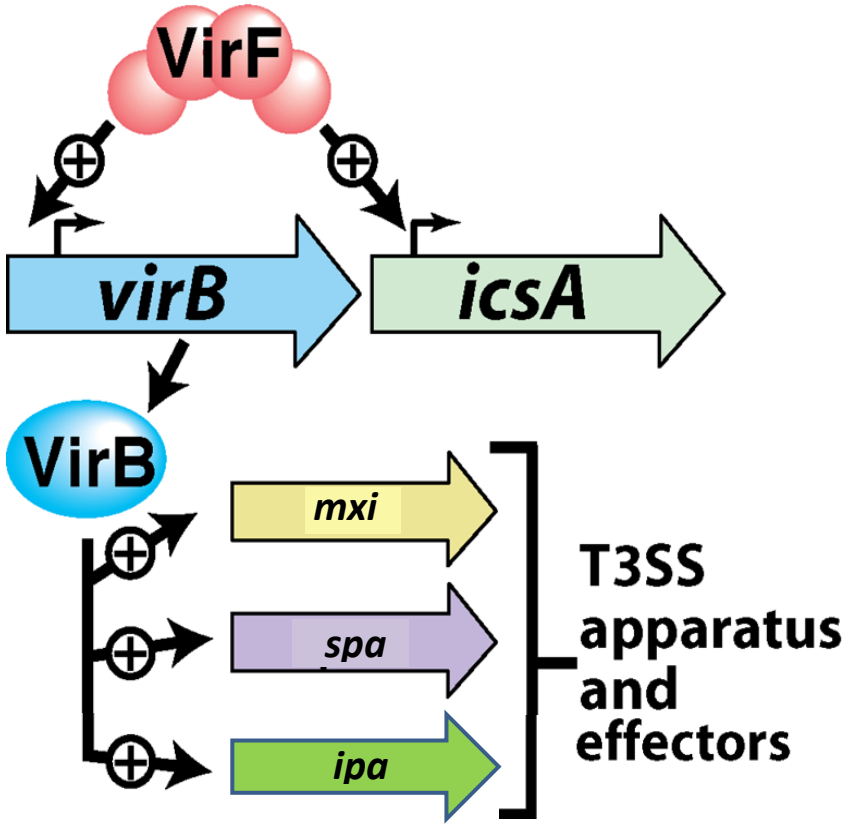


Fig. 3. VirF regulatory cascade. VirF activates transcription of the *virB* and *icsA* operons (148). VirB in turn activates expression of genes in *mxi*, *spa* and *ipa* operons that are required for the formation of T3SS machinery, invasion and cell-cell spread of *Shigella* into host epithelial cells (45-49). (Figure courtesy of Dr. Susan Egan).

temperatures below 30^{o}C, low osmolarity, and low pH by binding at these promoter regions within intrinsically curved regions (162-164). H-NS binds at two regions at *virF* promoter region (centered at -250 and -1 relative to TSS), three regions at *icsA* promoter region (centered at +25, +115 and +330 relative to TSS), and one region at *virB* promoter region (-20 to +20 TSS) (149, 162, 163, 165). H-NS binding sites at these promoter regions overlap binding site for RNA polymerase and therefore probably inhibit transcription initiation by preventing DNA binding by RNA polymerase (149). At 37^{o}C, a change in DNA structure results in the release of H-NS from DNA and thereby facilitates expression from these promoters (161, 162, 164, 166). Expression of VirF (during the stationary phase of the growth), IcsA and VirB (during all growth phases) are positively regulated by a protein called integration host factor (IHF) which binds at a 13-bp consensus sequence near the start of open reading frames (167). In case of VirF the positive regulation by IHF is confined to stationary phase of the bacterial growth, while for IcsA and VirB the positive regulation by IHF is observed irrespective of the bacterial growth phase (167).

VirF activates transcription of *icsA* and *virB* genes by directly binding at their respective promoter regions. DNase footprinting experiments showed four VirF binding sites (each ~40 to 60 bp in length) centered at -3, +62, +160 and +245 at *icsA* promoter region (165). At *virB* promoter region VirF was found to interact directly in a region spanning -17 to -105 (149). VirF activation of *virB* expression is enhanced by the presence of DNA superhelicity at the promoter region (168)

VirF is required for *Shigella* pathogenicity. The first step in the pathogenesis of *Shigella* is invasion into colonic and rectal epithelial cells. After invasion, *Shigella* replicates and spreads from cell-to-cell within the colonic and rectal epithelium. In addition to direct damage by

Shigella, host inflammatory responses also contribute to the epithelial cell damage leading to shigellosis (45-48, 165, 169-171).

VirF activates the expression of *icsA* which encodes an outer membrane protein IcsA (VirG) that helps in intracellular movement of the pathogen by mediating the actin based motility (172). IcsA gets localized at one pole of the bacterium and interacts with the host proteins N-WASP and vinculin. IcsA interaction with N-WASP and vinculin stimulates Arp2/3 which mediates actin polymerization (173-176) and thereby helps in movement.

VirF activates the expression of *virB* which encodes VirB, which in turn activates the expression of many virulence-associated operons, such as, *mxi*, *spa* and *ipa* operon (148, 177, 178). Genes in the *mxi* and *spa* operons encode the T3SS machinery through which effectors are released into host cells (52). Another AraC family protein MxiE is encoded from *mxi* operon and further activates the expression of many effectors involved in post-invasion steps of infection (52).

Genes in the *ipa* operon encode effector proteins (IcsB, IpaA, IpaB, IpaC and IpaD) which translocate directly into host cells (47, 179-182). IcsB, prevents the triggering of host autophagy, a process that can be used by host cells to export invading bacteria to lysosomes for degradation (183, 184). IcsB binds to outer membrane protein IcsA and competitively inhibits its interaction with the host autophagy protein Atg5, and thereby prevents the uptake of *Shigella* by autophagosome (183). In addition to interaction with Atg5, IcsB interacts with cholesterol and this interaction interferes with the formation of autophagosome (184). Another effector IpaB has major roles in regulation of secretion of other effectors, formation of pores in the host cell membrane, lysis of phagosomes and macrophage apoptosis through interactions with other

effector proteins (47, 179-182). IpaB interacts with IpaD in the secretion apparatus of the bacterial envelope and regulates the secretion of other Ipa proteins (182). IpaB interacts with IpaC and forms a hydrophobic complex that penetrates into the host cell membrane to form pores (179). Furthermore, IpaC of this complex integrates into the host cell membrane and stimulates Cdc42 (GTPase) that promotes membrane ruffles to facilitate the intake of bacteria into the host cells (180). The IpaB-IpaC complex is involved in lysing of phagosomes in macrophages and epithelial cells and thereby allows bacteria to gain free access to the host cytosol (47). IpaB-IpaC complex lyse the double cellular membranes between two adjacent epithelial cells leading to the movement of bacteria from cell to cell. When secreted into epithelial cells, IpaB interacts with pro-apoptotic capase-1 and triggers macrophage apoptosis (181). Apoptosis of macrophages in turn activates pro-inflammatory cytokines that induce the inflammatory responses leading to bacillary dysentery (181).

Given that VirF is required for VirB expression and that VirB activates the expression of multiple virulence-associated genes that are required for the earliest steps in the successful invasion of host cells by *Shigella*, VirF is considered the master regulator of *Shigella* virulence and works as an upstream regulator of all the virulence-associated genes (45-49). Further, mutants of *Shigella* that do not express VirF are avirulent (185, 186).

Similarities between virulence regulators Rns and VirF. Rns and VirF are virulence regulators in ETEC and *Shigella* respectively. Like many AraC family proteins both of them have structural domains NTD and DBD. Both Rns and VirF are fairly closely related to each other and share an amino acid sequence identity of 36%. Unlike the common situation in AraC family where relatedness is limited to DBD, Rns and VirF shares sequence similarities throughout their length. Both of them have substantially low G + C content (28%) compared to

the overall G + C content of their respective pathogens (50%). This, combined with their high sequence identity suggests that they may have a common ancestor outside their respective pathogens and were probably acquired by these pathogens by some mechanism (36, 187). Although Rns and VirF activate different sets of genes, it was found that Rns could activate the gens regulated by VirF and therefore could functionally substitute for VirF (188). In addition, CfaR which is ~95% identical to Rns could functionally substitute for VirF (188). Both Rns and VirF are sensitive to changes in temperature and activate transcription at 37^{o}C from promoter regions that are subject to negative regulation by H-NS at low temperatures (41, 149).

The AraC family activator RhaR

RhaR is an AraC family protein that activates the transcription of *rhaSR* operon that encodes both RhaS and RhaR proteins (102, 189, 190). RhaS in turn activates the transcription of *rhaBAD* and *rhaT* operons (189, 191, 192) (Fig. 4). *rhaT* encodes a transporter protein that is involved in the uptake of L-rhamnose into the *E. coli* cells (193). The *rhaBAD* operon encodes proteins that are required for sugar L-rhamnose catabolic breakdown (*rhaB* encodes L-rhamnulokinase that converts L-rhamnulose into L-rhamnuose 1-phospate (194); *rhaA* encodes an isomerase that converts L-rhamnose to L-rhamnulose (195); *rhaD* encodes an aldolase that breaks down L-rhamnuose 1-phospate into dihydroxy acetone phosphate and L- lactaldehyde (196). Transcription activation by RhaR requires the presence of effector L-rhamnose. In the absence of rhamnose, RhaR can bind to the DNA but activates transcription of genes encoding RhaS and RhaR only to low basal levels. But when the RhaR effector L- rhamnose is available, RhaR can activate transcription to a 5-7 fold higher level relative to the absence of rhamnose (102, 189, 190, 197, 198).

Previously, it was found that the DNA binding domain (DBD) of RhaR, when expressed alone, can bind to the DNA but does not respond to sugar L-rhamnose, suggesting that L-rhamnose binding site is in the NTD (76). It was found that DBD in the full length RhaR activates transcription through interactions with RNAP σ70 and C-terminal domain (CTD) of the RNAP α-subunit (81, 199). The separation of L-rhamnose binding and transcription activation activities between the two domains suggests that there must be some form of communication between NTD and DBD to correlate transcription activation to L-rhamnose binding. Therefore, we hypothesized that the signal of L-rhamnose binding in the NTD is somehow communicated to the DBD, and thus causing structural or conformational changes required by the DBD to activate transcription (presumably by binding to the DNA with higher affinity and contacting RNAP). Previous work from our lab (by Ana Kolin) has indicated that the linker that connects the RhaR NTD to DBD does not play any crucial role in the activation, suggesting that L-rhamnose dependent allosteric signals are not mediated through the linker region (74). To identify the mechanism of RhaR L-rhamnose response, she further tested the hypothesis that RhaR N-terminal residues have a role in L-rhamnose response that is similar to N-terminal residues of AraC that play a role in AraC L-arabinose response (as described by light-switch mechanism above). She made a series of RhaR N-terminal truncations and found that the deletion of residues up to the first 34 amino acids did not have any role in RhaR L-rhamnose response (200). N-terminal truncations extending after the first 34 amino acids (Δ40 or more) resulted in decreased activation levels along with decreased protein levels, and thus restricted from drawing any conclusions (200). However, alignment of AraC with the RhaR showed that AraC Arm aligns with the RhaR residues 35-43. Further, structural prediction model I-TASSER has predicted that RhaR residues from 35-45 may form a loop that is positioned to contact residues in the DBD.

These observations led us to hypothesize that RhaR residues 35-45 may have a role as analogous to the role of AraC Arm residues.

Targeting of AraC family proteins to develop novel anti-bacterial agents. Given that AraC family proteins are required for the expression of virulence genes in many human pathogens, they are considered as potential targets for novel antibacterial agents. To develop novel antibacterials, AraC family proteins can be targeted in two fundamentally different paths: 1) high-through put screening of large-libraries of molecules against AraC family proteins 2) designing inhibitors using a rational drug design approach (201). A successful rational drug design approach to target AraC family proteins would need either structure of the target or a better understanding of the molecular interactions involved in transcription activation by these proteins. Using high-throughput screening, Dr. Susan Egan lab recently identified an AraC family inhibitor, SE-1 (previously called OSSL_051168), which selectively inhibited two AraC family activators RhaS and RhaR (activators of L-rhamnose catabolic operons in *E. coli*), despite their limited sequence similarity with each other (202). We further found that SE-1 inhibited RhaS and the RhaS DNA-binding domain (RhaS-DBD) to the same extent, and that it blocked DNA binding by RhaS and RhaR (202). Furthermore, SE-1 had little or no impact on DNA binding by the non-AraC family proteins LacI and CRP, indicating selectivity for at least the tested AraC family proteins (202). Taken together, our results support the hypothesis that SE-1 acts on the RhaS and RhaR DBDs - the structurally conserved domain in AraC family proteins. These observations lead us to propose the hypothesis that this compound might selectively inhibit additional AraC family activators.

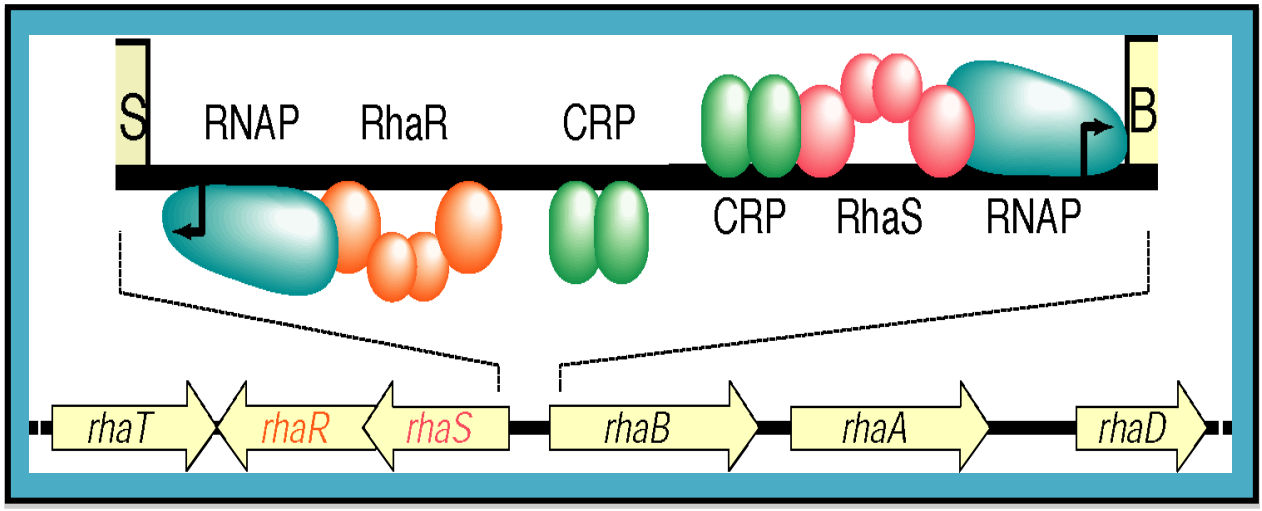


Fig. 4. E. coli L-rhamnose regulon. The relative positions of genes in the L-rhamnose operon are presented as arrows. DNA was shown as a black horizontal line. Proteins that bind at the *rhaBAD* and *rhaSR* promoters and their relative positions are shown on the DNA (Figure courtesy of Dr. Susan Egan).

Research goals

The overall goals of my research are to better understand the transcription activation mechanisms of Rns and RhaR, and to test a small molecule for inhibition of Rns and VirF. The overall goals can be divided into the following four subgoals.

Mechanism of transcription activation in ETEC virulence activator Rns: The molecular mechanism of Rns activation of fimbrial genes is not clearly understood. Evidence suggests that Rns-DBD alone is not capable of activating transcription and requires the presence of regulatory domain (NTD) (156). This is supported by the deletion analysis of NTD regions and identification of Rns-NTD variants I14T and N16D at the N-terminus of the protein that strongly reduced activation at *rns* promoter (156). We proposed that transcription activation mechanism in Rns involves contacts between residues in the Rns-NTD and Rns-DBD. In the absence of crystal structure of Rns, we used the only available crystal structure of ToxT (79) (the only available structure in AraC family virulence activators) as our model. ToxT crystal structure solved with fatty acid effector bound in the effector binding pocket showed that β2-strand of NTD and several residues in the DBD are positioned to make inter-domain interactions (79). Rns shares ~20% identity and 46% similarity with ToxT and structural prediction model (I-TASSER) has predicted a structure for Rns that is very similar to ToxT (79, 157). I-TASSER predicted multiple contacts between a region in the Rns-NTD and several residues in the Rns-DBD. The region in the Rns-NTD that is analogous to the ToxT-NTD region that contacts the DBD is comprised of residues I12-M18 (79). We referred this region of residues in Rns-NTD as the regulatory site2 (RS2). The region in the Rns-DBD that is predicted to be contacted by the residues in the RS2 region includes several residues (Q211, S215, K216, L219, K249-G252 and

Y261), which we refer to as the allosteric site (AS2). Residues in both RS2 and AS2 regions also align with ToxT regions involved in making inter-domain interactions.

In order to understand the roles RS2 and AS2 regions may have in transcription activation of Rns, we have mutagenized residues in RS2 and AS2 regions and studied their effect in Rns dependent LacZ expression from *cfaA-lacZ*$^{\Delta 60}$ (Wild-type Rns activates lacZ expression from this fusion) and *nlpA*^{cfaA}*-lacZ* fusion (Wild-type Rns represses lacZ expression from this fusion). We identified multiple variants in specific RS2 region residues (N15, N16 and I17) and AS2 region residues (K216, H250, Y251 and G252) that conferred decreased Rns-dependent expression of *cfaA-lacZ*$^{\Delta 60}$ compared to the wild type, suggesting that specific residues in these regions are required for Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression. Further analysis of variants for repression of Rns-dependent LacZ expression from *nlpA*^{cfaA}*-lacZ* fusion resulted in distinguishing residues that are required for DNA binding from transcription activation (In this context, we refer transcription activation as an aggregate of processes such as contacts with RNA polymerase, alteration of promoter DNA conformation and relief of repression). Our results suggested that residues N15, N16, I17, K216, Y251 and G252 are likely required for transcription activation and residue H250 is likely required for both transcription activation and DNA binding.

Purification and properties of Rns. The second goal of this study was purification and biochemical characterization of Rns protein. I purified GB1 tagged Rns-DBD and MBP tagged full-length Rns using affinity purification and tested for their ability to bind their specific DNA in electrophoretic mobility shift assays (EMSAs). Behavior of MBP-Rns *in vitro* was further tested in gel filtration analysis, thermo flour analysis and dynamic light scattering analysis. The purified protein was used to raise antibodies against Rns for use in Western blotting.

Small molecule inhibitor of AraC family proteins Rns and VirF. The third goal of my research was to test a small molecule compound, SE-1 (previously OSSL_051168), which our lab had identified as an effective inhibitor of the AraC family proteins RhaS and RhaR (202), for its ability to inhibit virulence activators Rns and VirF. Preliminary work by Jeff Skredenske suggested that SE-1 inhibits Rns (unpublished) and VirF (203) activation of LacZ expression in *E. coli* cell-based reporter gene assays . I further found that SE-1 effectively inhibit Rns and VirF DNA binding using *in vitro* DNA binding assays. SE-1 was tested for its ability to inhibit VirF activation in *Shigella* using reporter gene assays. Analysis of mRNA levels for direct and indirect targets of VirF in the presence of SE-1was performed by real-time quantitative reverse transcription PCR (qRT-PCR) and showed decreased expression of all of the VirF dependent virulence genes tested. Invasion assays were performed (done by Ichie Osaka) to test the ability of SE-1 in inhibition of *Shigella* invasion of epithelial cells in tissue culture cells and found that SE-1 inhibits invasion by *Shigella*.

Role of RhaR Arm in L-rhamnose response. My fourth goal of this study was to understand the role of Arm residues (35-45) in RhaR in transmitting L-rhamnose signal from RhaR-NTD to RhaR-DBD. Evidence suggests that L-rhamnose binds at the NTD of RhaR, while transcription activation is performed by RhaR-DBD. Therefore, one hypothesis is that signal is transmitted from RhaR-NTD to RhaR-DBD in response to L-rhamnose binding. Here, I have tested whether that signal is transmitted through Arm residues (35-45) in RhaR that align with Arm residues in AraC. I mutagenized residues in the RhaR Arm region by site-directed random mutagenesis and isolated a total of 46 unique variants. Our results suggested that residue L35 in the Arm may be involved in making inhibitory inter-domain interactions in the (-) of rhamnose.

Chapter 2

MATERIALS AND METHODS

Growth conditions. *Escherichia coli* strains were grown in tryptone-yeast extract (TY) broth (0.8% Difco tryptone, 0.5% Difco yeast extract, and 0.5% NaCl, [pH 7.0]). Difco nutrient agar (1.5% agar) (BD, Cockeysville, MD) was used routinely to grow *E. coli* strains on solid medium. To test for *lacZ* expression by Rns and RhaR on solid media, Difco nutrient agar was supplemented with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside, 40 μg/ml) and X-gal with L-rhamnose (0.2%), respectively. All liquid cultures were grown at 37°C with shaking, except for cultures carrying the temperature sensitive plasmid pAH69 pAH69, (204) which were grown at 30°C. β-galactosidase assays in *E. coli* strains that express RhaR were grown in the presence or absence of 0.4% L-rhamnose in a MOPS [3-(*N*-morpholino)propanesulfonic acid]-buffered minimal medium (40 mM MOPS, 4 mM Tricine, 10 μM $FeSO_4$, 9.5 mM NH_4Cl, 0.276 mM K_2SO_4, 0.5 μM $CaCl_2$, 0.528 mM $MgCl_2$, 50 mM NaCl, 3 nM Na_2MoO_4, 0.4 μM H_3BO_3, 30 nM $CoCl_2$, 10 nM $CuSO_4$, 80 nM $MnCl_2$, 10 nM $ZnSO_4$, 1.32 mM K_2HPO_4, 10 mM $NaHCO_3$, 0.002% casaminoacids, 0.002% thiamine) with a carbon source of 0.4% glycerol.

Cultures of *Shigella* strains were grown in tryptic soy broth (TSB, pH 7.0) (BD), Luria-Bertani broth (1% tryptone, 0.5% yeast extract and 1% NaCl) or on tryptic soy agar (TSA, 1.5% agar) (BD) containing congo red dye (0.025%). All *Shigella* strains were picked as red colonies from TSA plates containing Congo red. All *Shigella* strains were grown at 37^{o}C with shaking, unless otherwise specified. Appropriate antibiotics (ampicillin, 200 μg/ml; chloramphenicol, 30 μg/ml; tetracycline, 20 μg/ml; and spectinomycin, 100 μg/ml) were used as indicated.

General Methods. Primers used in this study were synthesized by Eurofins MWG Operon (Huntsville, AL). PCR was performed using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). Amplified DNA fragments were cleaned up using either E.Z.N.A. Cycle Pure (Omega Bio-tek, Norcross, GA) or IBI Gel/PCR products extraction kits (IBI Scientific, Peosta,

IA). Standard methods were followed for restriction endonuclease digestion, ligation, and transformation. Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). Plasmid isolation from bacteria was performed using the IBI High-Speed Plasmid Mini kit (MIDSCI, St. Louis, MO). DNA sequencing was performed to obtain sequences of for both strands of the entire cloned region for all cloned and mutagenized DNA fragments at the NorthWestern University Genomics Core (Chicago, IL).

Bacterial strains, plasmids and oligos. Bacterial strains, plasmids and oligos used in this study are shown in Tables 1, 2, 3 respectively.

Construction of Rns and Rns-DBD plasmids for *in vivo* experiments. Rns and Rns-DBD were expressed from *lac* promoter in a low copy number plasmid pHG165 (Amp^r) (205). To make these plasmids, *rns* was first subcloned from pEU2030 (a pUC18 based high copy number plasmid, ampicillin resistant) into pSU18 (a low copy number plasmid, chloramphenicol resistant) to make pSU18/*rns*. The subcloning method consisted of PCR amplification of the *rns* using primers 2963 and 2964. The forward primer 2963 had Shine-Dalgarno sequence (part of the ribosome binding site). The PCR products were then digested and ligated into pSU18 plasmid at *Eco*RI and *Hind*III sites. The plasmid pSU18/*rns* was then used as template to amplify the regions encoding Rns and Rns-DBD (residues 157-265) including the Shine-Dalgarno sequence. The PCR products were digested and then ligated into pHG165 such that the *lac* promoter of the pHG165 would drive the transcription, but translation initiation site was part of the insert. This resulted in plasmids pHG165/*rns* and pHG165/*rns* DBD (residues 157-265).

Table 1. Bacterial strains used in this study.

Strain	Genotype	Source or Reference
MC4100	F$^-$ *araD139* Δ (*argF-lac*) *U169 rpsL150 relA1 flhD5301 deoC1 ptsF25 rbsR*	(206)
KS1000	F'*lacI*^{q} *lac*$^+$ *pro*$^+$*/ ara* Δ*(lac-pro)* Δ*(tsp)* Δ*(prc)::Kan*^{R} *eda51::Tn10 tet*^{R} *gyrA(Nal*^{R}*) rpoB thi-1 argE(am)*	(207)
AF51	MC4100 *attB*HK022::pCFAILac10	(156)
AB97	MC4100 λMAD102 [Φ (*virB-lacZ*)] -259 to +308 relative to transcription start site	A. Maurelli
SME1362	DH5α	Laboratory collection
ECL116	F$^-$ Δ*lacU169 endA hsdR thi*	Laboratory collection
AJ22 (SME3989)	DE*(argF-lac)169uidA4(del)::pir-116recA1 rpoS396(Am) endA9(del-ins)::FRT rph-1 hsdR514 creC510 robA1*	George Munson
SME1048	ECL116 *recA*::*cat*	Laboratory collection
SG22166	MC4100 *malP*::*lacI*^{q} *ftsH1*(ts) *zgj*::Tn*10*	
SME2525	ECL116 λΦ(*rhaS-lacZ*)Δ128 Δ*rhaSR*::*kan recA::cat*	Laboratory collection
SME3110	ECL116 λΦ(*rhaS-lacZ*)Δ128 Δ*rhaSR::kan*	Laboratory collection
SME3160	ECL116 λΦ(*rhaS-lacZ*)Δ85 Δ*rhaSR::kan recA::cat*	Laboratory collection
SME3358	ECL116 Δ*rhaS* λ Φ (P*hts-lacZ*) *recA*::*cat*	(202)
SME3359	SME3358 + pHG165/*lacI*	This study
SME3613(AP76)	MC4100 *attB*HK022::pNLPALac1	(156)
SME 3670	AF51+ pHG165/*rns*	This study
SME 3676	AF51 + pSU18/*rns*	This study
SME 3677	SME 3110 + pSU18/*rhaR*	This study
SME3740	SME 2525 + pHG165/*rhaR* L35V	This study
SME3741	SME 2525 + pHG165/*rhaR* L35K	This study
SME3742	SME 2525 + pHG165/*rhaR* L35Y	This study
SME3743	SME 2525 + pHG165/*rhaR* L35K	This study
SME3744	SME 2525 + pHG165/*rhaR* L35I	This study
SME3745	SME 2525 + pHG165/*rhaR* L35G	This study
SME3747	SME 2525 + pHG165/*rhaR* K36V	This study
SME3749	AF51 *recA::cat*	This study
SME3752	AF51+ pHG165/*rns* (117-265)	This study
SME3779	SME 3754 + pHG165/*rns* (157-265)	This study
SME3780	SME 3754 + pHG165	This study
SME3781	SME2525 + pHG165/*rhaR* L38V	This study
SME3782	SME2525 + pHG165/*rhaR* L38S	This study
SME3783	SME2525 + pHG165/*rhaR* L38G	This study
SME3784	SME2525 + pHG165/*rhaR* L38R	This study

SME3785	SME2525 + pHG165/*rhaR* L38W	This study
SME3786	SME2525 + pHG165/*rhaR* L37I	This study
SME3787	SME2525 + pHG165/*rhaR* L37R	This study
SME3750	BL21(DE3) DnaY cells (Km^r)	Laboratory collection
SME3805	Rosetta2(DE3)	Novagen
SME3817	SME3750 + pDZ1/*rns*	This study
SME3818	SME3750 + pDZ1/*rns* NTD	This study
SME3819	SME3750 + pDZ3/*rns*	This study
SME3820	SME3750 + pDZ3/*rns* NTD	This study
SME3863	SME3750 + pDZ3/*rns* DBD	This study
SME3864	SME3805 + pDZ1/*rns*	This study
SME3865	SME3805 + pDZ1/*rns* NTD	This study
SME3952	SME2525 + pHG165/*rhaR* K36A	This study
SME3953	SME2525 + pHG165/*rhaR* K36R	This study
SME3954	SME2525 + pHG165/*rhaR* L37C	This study
SME3955	SME2525 + pHG165/*rhaR* L37S	This study
SME3956	SME2525 + pHG165/*rhaR* L37G	This study
SME3957	SME2525 + pHG165/*rhaR* L37M	This study
SME3958	SME2525 + pHG165/*rhaR* L37A	This study
SME3961	SME2525 + pHG165/*rhaR* L35D	This study
SME3963	SME3160 + pHG165/*rhaR* L35Y	This study
SME3964	SME3160 + pHG165/*rhaR* L35V	This study
SME3965	SME3160 + pHG165/*rhaR* L35K	This study
SME3966	SME3160 + pHG165/*rhaR* L35I	This study
SME3967	SME3160 + pHG165/*rhaR* L35G	This study
SME3968	SME3160 + pHG165/*rhaR* L35D	This study
SME3969	SME3160 + pHG165/*rhaR* K36R	This study
SME3970	SME3160 + pHG165/*rhaR* K36A	This study
SME3971	SME3160 + pHG165/*rhaR* K36V	This study
SME3972	SME3160 + pHG165/*rhaR* L37I	This study
SME3973	SME3160 + pHG165/*rhaR* L37R	This study
SME3974	SME3160 + pHG165/*rhaR* L37M	This study
SME3975	SME3160 + pHG165/*rhaR* L37S	This study
SME3976	SME3160 + pHG165/*rhaR* L37A	This study
SME3977	SME3160 + pHG165/*rhaR* L37C	This study
SME3978	SME3160 + pHG165/*rhaR* L37G	This study
SME3979	SME3160 + pHG165/*rhaR* L38V	This study
SME3980	SME3160 + pHG165/*rhaR* L38R	This study
SME3981	SME3160 + pHG165/*rhaR* L38G	This study
SME3982	SME3160 + pHG165/*rhaR* L38S	This study
SME3983	SME3160 + pHG165/*rhaR* L38W	This study
SME3985	SME3749 + pHG165/*rns*	This study
SME3991	SME3160 + pHG165/*rhaR* K39A	This study
SME3992	SME3160 + pHG165/*rhaR* K39V	This study
SME3993	SME3160 + pHG165/*rhaR* K39P	This study

SME3994	SME3160 + pHG165/*rhaR* K39E	This study
SME3995	SME3160 + pHG165/*rhaR* K39T	This study
SME3996	SME3160 + pHG165/*rhaR* K39S	This study
SME3997	SME3160 + pHG165/*rhaR* K39P	This study
SME3998	SME3160 + pHG165/*rhaR* D40E	This study
SME3999	SME3160 + pHG165/*rhaR* D40G	This study
SME4000	SME3160 + pHG165/*rhaR* D41S	This study
SME4001	SME3160 + pHG165/*rhaR* D41A	This study
SME4002	SME3160 + pHG165/*rhaR* D41V	This study
SME4003	SME3160 + pHG165/*rhaR* F42M	This study
SME4004	SME3160 + pHG165/*rhaR* F42I	This study
SME4005	SME3160 + pHG165/*rhaR* F42D	This study
SME4006	SME3160 + pHG165/*rhaR* F42N	This study
SME4007	SME3160 + pHG165/*rhaR* F42P	This study
SME4008	SME3160 + pHG165/*rhaR* F43V	This study
SME4009	SME3160 + pHG165/*rhaR* F43C	This study
SME4010	SME3160 + pHG165/*rhaR* F43R	This study
SME4011	SME3160 + pHG165/*rhaR* F43A	This study
SME4012	SME3160 + pHG165/*rhaR* F43G	This study
SME4013	SME3160 + pHG165/*rhaR* A44G	This study
SME4014	SME3160 + pHG165/*rhaR* A44E	This study
SME4015	SME3160 + pHG165/*rhaR* S45V	This study
SME4016	SME3160 + pHG165/*rhaR* S45G	This study
SME4017	SME3160 + pHG165/*rhaR* S45T	This study
SME4019	SME1362 + pDZ1/*rns*	This study
SME4020	SME1362 + pSE295/*rns*	This study
SME4021	ECL116 λMAD102 [Φ (*virB-lacZ*)] -259 to +308 relative to transcription start site	(202)
SME4143	KS1000 + pMBP-Rns1	This study
SME4147	KS1000 + pMBP-Rns1 + pRARE2	This study
SME4149	SME1362 + pSE290/ rns^{co}	This study
SME4150	SME1362 + pSE295/rns^{co}	This study
SME4219	SME3160 + pHG165/*rhaR* F42L	This study
SME4220	SME2525 + pHG165/*rhaR* F43S	This study
SME4240	SME4021 *malP*::$lacI^q$ *zhc-511*:: Tn10	This study
SME4259	SME4240 *recA*::*cat*	This study
SME4319	KS1000 + pMAL/*virF*	This study
SME4324	MC4100 *attB*HK022::pSE298	This study
SME4335	SME4324 *recA::cat*	This study
SME4382	SME4259 + pHG165/*virF*	This study
SME4336	SME4335 + pHG165/*rns*	This study
SME4337	SME4335 + pHG165	This study
SME4344	SME4335 + pHG165/*rns* I12R	This study
SME4345	SME4335 + pHG165/*rns* I12S	This study
SME4346	SME4335 + pHG165/*rns* I12G	This study

SME4347	SME4335 + pHG165/*rns* I12E	This study
SME4348	SME4335 + pHG165/*rns* I12W	This study
SME4349	SME4335 + pHG165/*rns* I14W	This study
SME4350	SME4335 + pHG165/*rns* N15L	This study
SME4351	SME4335 + pHG165/*rns* N15D	This study
SME4352	SME4335 + pHG165/*rns* N15G	This study
SME4353	SME4335 + pHG165/*rns* N16C	This study
SME4354	SME4335 + pHG165/*rns* N16D	This study
SME4355	SME4335 + pHG165/*rns* N16G	This study
SME4356	SME4335 + pHG165/*rns* K249R	This study
SME4357	SME4335 + pHG165/*rns* K249Y	This study
SME4358	SME4335 + pHG165/*rns* K249A	This study
SME4359	SME4335 + pHG165/*rns* K249E	This study
SME4360	SME4335 + pHG165/*rns* K249N	This study
SME4361	SME4335 + pHG165/*rns* H250R	This study
SME4362	SME4335 + pHG165/*rns* H250I	This study
SME4363	SME4335 + pHG165/*rns* H250N	This study
SME4364	SME4335 + pHG165/*rns* H250T	This study
SME4365	SME4335 + pHG165/*rns* H250V	This study
SME4366	SME4335 + pHG165/*rns* H251F	This study
SME4367	SME4335 + pHG165/*rns* H251T	This study
SME4368	SME4335 + pHG165/*rns* H251L	This study
SME4369	SME4335 + pHG165/*rns* H251V	This study
SME4370	SME4335 + pHG165/*rns* G252N	This study
SME4371	SME4335 + pHG165/*rns* G252Q	This study
SME4372	SME4335 + pHG165/*rns* G252R	This study
SME4373	SME4335 + pHG165/*rns* G252S	This study
SME4374	SME4335 + pHG165/*rns* G252A	This study
SME4375	SME4335 + pHG165/*rns* Y261G	This study
SME4376	SME4335 + pHG165/*rns* Y261E	This study
SME4377	SME4335 + pHG165/*rns* Y261M	This study
SME4378	SME4335 + pHG165/*rns* Y261A	This study
SME4379	SME4335 + pHG165/*rns* Y261R	This study
SME4380	SME4335 + pHG165/*rns* Y261C	This study
SME4382	SME4259 + pHG165/*virF*	This study
SME4383	SME4259 + pHG165/*rns*	This study
SME4436	SME3989 + pSE299	This study
SME4437	SME4335 + pHG165/*rns* I17V	This study
SME4438	SME4335 + pHG165/*rns* I17C	This study
SME4439	SME4335 + pHG165/*rns* I17P	This study
SME4440	SME4335 + pHG165/*rns* I17R	This study
SME4441	SME4335 + pHG165/*rns* I17N	This study
SME4442	SME4335 + pHG165/*rns* I17Y	This study
SME4443	SME4335 + pHG165/*rns* I17S	This study
SME4444	SME4335 + pHG165/*rns* M18R	This study

SME4445	SME4335 + pHG165/*rns* M18P	This study
SME4446	SME4335 + pHG165/*rns* M18G	This study
SME4447	SME4335 + pHG165/*rns* Q211G	This study
SME4450	SME4335 + pHG165/*rns* Q211S	This study
SME4451	SME4335 + pHG165/*rns* Q211L	This study
SME4452	SME4335 + pHG165/*rns* Q211N	This study
SME4453	SME4335 + pHG165/*rns* S215V	This study
SME4455	SME4335 + pHG165/*rns* S215G	This study
SME4457	SME4335 + pHG165/*rns* S215K	This study
SME4458	SME4335 + pHG165/*rns* S215D	This study
SME4459	SME4335 + pHG165/*rns* K216P	This study
SME4460	SME4335 + pHG165/*rns* K216L	This study
SME4461	SME4335 + pHG165/*rns* L219M	This study
SME4463	SME4335 + pHG165/*rns* L219H	This study
SME4464	SME4335 + pHG165/*rns* L219P	This study
SME4465	SME4335 + pHG165/*rns* L219A	This study
SME4466	SME4335 + pHG165/*rns* L219Y	This stud
SME4467	SME4335 + pHG165/*rns* L219K	This study
SME4468	SME4335 + pHG165/*rns* L219I	This study
SME4471	SME4335 + pHG165/*rns* K13S	This study
SME4472	SME4335 + pHG165/*rns* K13G	This study
SME4473	SME4335 + pHG165/*rns* K13C	This study
SME4474	MC4100 *attB*HK022::pSE299	This study
SME4475	SME4474 + pHG165/*rns* I12R	This study
SME4476	SME4474 + pHG165/*rns* I12S	This study
SME4477	SME4474 + pHG165/*rns* I12G	This study
SME4478	SME4474 + pHG165/*rns* I12E	This study
SME4479	SME4474 + pHG165/*rns* I12W	This study
SME4480	SME4474 + pHG165/*rns* K13S	This study
SME4481	SME4474 + pHG165/*rns* K13G	This study
SME4482	SME4474 + pHG165/*rns* K13C	This study
SME4483	SME4474 + pHG165/*rns* I14W	This study
SME4484	SME4474 + pHG165/*rns* N15L	This study
SME4485	SME4474 + pHG165/*rns* N15D	This study
SME4486	SME4474 + pHG165/*rns* N15G	This study
SME4487	SME4474 + pHG165/*rns* N16C	This study
SME4488	SME4474 + pHG165/*rns* N16D	This study
SME4489	SME4474 + pHG165/*rns* N16G	This study
SME4490	SME4474 + pHG165/*rns* I17N	This study
SME4491	SME4474 + pHG165/*rns* I17R	This study
SME4492	SME4474 + pHG165/*rns* I17S	This study
SME4493	SME4474 + pHG165/*rns* I17V	This study
SME4494	SME4474 + pHG165/*rns* I17Y	This study
SME4495	SME4474 + pHG165/*rns* I17C	This study
SME4496	SME4474 + pHG165/*rns* I17P	This study

SME4497	SME4474 + pHG165/*rns* M18P	This study
SME4498	SME4474 + pHG165/*rns* M18R	This study
SME4499	SME4474 + pHG165/*rns* M18G	This study
SME4501	SME4474 + pHG165/*rns* Q211S	This study
SME4502	SME4474 + pHG165/*rns* Q211G	This study
SME4503	SME4474 + pHG165/*rns* Q211L	This study
SME4504	SME4474 + pHG165/*rns* Q211N	This study
SME4505	SME4474 + pHG165/*rns* S215V	This study
SME4507	SME4474 + pHG165/*rns* S215G	This study
SME4508	SME4474 + pHG165/*rns* S215D	This study
SME4509	SME4474 + pHG165/*rns* S215K	This study
SME4510	SME4474 + pHG165/*rns* K216P	This study
SME4511	SME4474 + pHG165/*rns* K216L	This study
SME4512	SME4474 + pHG165/*rns* L219M	This study
SME4513	SME4474 + pHG165/*rns* L219I	This study
SME4514	SME4474 + pHG165/*rns* L219P	This study
SME4515	SME4474 + pHG165/*rns* L219A	This study
SME4516	SME4474 + pHG165/*rns* L219Y	This study
SME4517	SME4474 + pHG165/*rns* L219K	This study
SME4518	SME4474 + pHG165/*rns* L219H	This study
SME4519	SME4474 + pHG165/*rns*	This study
SME4520	SME4474 + pHG165	This study
SME4521	SME4474 + pHG165/*rns* K249R	This study
SME4522	SME4474 + pHG165/*rns* K249Y	This study
SME4523	SME4474 + pHG165/*rns* K249A	This study
SME4524	SME4474 + pHG165/*rns* K249E	This study
SME4525	SME4474 + pHG165/*rns* K249N	This study
SME4526	SME4474 + pHG165/*rns* H250R	This study
SME4527	SME4474 + pHG165/*rns* H250I	This study
SME4528	SME4474 + pHG165/*rns* H250N	This study
SME4529	SME4474 + pHG165/*rns* H250T	This study
SME4530	SME4474 + pHG165/*rns* H250V	This study
SME4531	SME4474 + pHG165/*rns* Y251F	This study
SME4532	SME4474 + pHG165/*rns* Y251T	This study
SME4533	SME4474 + pHG165/*rns* Y251L	This study
SME4534	SME4474 + pHG165/*rns* Y251V	This study
SME4535	SME4474 + pHG165/*rns* G252N	This study
SME4536	SME4474 + pHG165/*rns* G252Q	This study
SME4537	SME4474 + pHG165/*rns* G252R	This study
SME4538	SME4474 + pHG165/*rns* G252S	This study
SME4539	SME4474 + pHG165/*rns* G252A	This study
SME4540	SME4474 + pHG165/*rns* Y261G	This study
SME4541	SME4474 + pHG165/*rns* Y261E	This study
SME4542	SME4474 + pHG165/*rns* Y261M	This study
SME4543	SME4474 + pHG165/*rns* Y261A	This study

SME4544	SME4474 + pHG165/*rns* Y261R	This study
SME4545	SME4474 + pHG165/*rns* Y261C	This study
SME4590	SME3989 + pSE298	This study
S.flexneri strains		
BS536	*S.flexneri virB-lacZ*	A. Maurelli
SME4331	*S.flexneri ipgD*	W. D Picking

Table 2. Plasmids used in this study.

Plasmid	Relevant characteristics	Source or Reference
pHG165	*lacZα rop* Amp^r	
pHG165/*virF*	*virF* expressed from pHG165 *lac* promoter, Amp^r	This Study
pHG165/*lacI*	*lacI* expressed from pHG165 *lac* promoter, Amp^r	(202)
pHG165/*rns*	*rns* expressed from pHG165 *lac* promoter, Amp^r	This study
pHG165/*rns* (117-265)	*rns* (117-265) expressed from pHG165 *lac* promoter, Amp^r	This study
pHG165/*rns* (157-265)	*rns* (157-265) expressed from pHG165 *lac* promoter, Amp^r	This study
pSU18	LacZα Rep(p15A), Cm^r	
pSU18/*rns*	*rns* expressed from pSU18 *lac* promoter, Cm^r	This study
pSU18/*rhaR*	*rhaR* expressed from pSU18 *lac* promoter, Cm^r	This study
pAH69	Temperature sensitive helper plasmid, Amp^r	(204)
pHKLac1	pRS550 *lacZYA* + pAH144 *pir* origin, $Spec^r$	(39)
pSE298	pHKLac1 $cfaA\text{-}lacZ^{\Delta 60}$, $Spec^r$	This study
pSE299	pHKLac1 $nlpA^{cfaA}\text{-}lacZ$, $Spec^r$	This study
pDZ1	T7 promoter, N terminal His6-GB1 tag, Amp^r	R. DeGuzman
pDZ3	T7 promoter, C terminal GB1-His6 tag, Amp^r	R. DeGuzman
pDZ1/*rns*	*rns* expressed from pDZ1 T7 promoter, Amp^r	This study
pDZ1/*rns* NTD	*rns* NTD expressed from pDZ1 T7 promoter, Amp^r	This study
pDZ3/*rns*	*rns* expressed from pDZ3 T7 promoter, Amp^r	This study
pDZ3/*rns* NTD	*rns* NTD expressed from pDZ3 T7 promoter, Amp^r	This study
pDZ3/*rns* DBD	*rns* DBD expressed from pDZ3 T7 promoter, Amp^r	This study
pSE290	$GB1^{basic}$ version of pDZ1, Amp^r	Laboratory collection
pSE295	pDZ3 with $GB1^{basic}$ + linker, Amp^r	Laboratory collection
pSE295/rns^{co}	rns^{co} expressed from pSE295 T7 promoter, Amp^r	This study
pSE290/rns^{co}	rns^{co} expressed from pSE290 T7 promoter, Amp^r	This study
pMAL-c2X	$lacI^q$ *malE-lacZα,* Amp^r	New England Biolabs
pMAL/*virF*	*virF* from pMAL-c2X, Amp^r	This study
pMAL/*rns*	*rns* from pMAL-c2X, Amp^r	This study
pMAL/*rns* NTD	*rns* NTD from pMAL-c2X, Amp^r	This study
pMAL/*rns* DBD	*rns* DBD from pMAL-c2X, Amp^r	This study
pRARE2	Provides seven rare tRNAs to supplement *rns* rare tRNA usage, Cm^r	Novagen
pMBP-Rns1	pMal-c2X *rns*, Amp^r	(39)

Table 3. Oligonucleotides used in this study.

Oligo No.	Oligonucleotide sequence 5'-3'	Use
2962	GCGGAATTCCTTTTGTAGGTATAAGATGGAC	Amplify *rns* and *rns*^{co} (upstream oligo), contains Shine-Dalgarno sequence
2963	GCGAAGCTTTTATCCACCTTTAAAATAAGTG	Amplify *rns* (downstream oligo)
2964	AAAGAATTCCTTTTGTAGGTATAAGATGACCTT GTTGGATGAA	Upstream oligo to construct Rns (117-265)
2965	GCGGAATTCCTTTTGTAGGTATAAGATGTCATC TGTGAGT-3'	Upstream oligo to construct Rns (157-265)
2980	TCTCTCTTCGCGCATCAGNNNAAACTTCTCAAA GATGAT	Site-directed random mutagenesis of *rhaR* L35 (middle downstream)
2981	TCTCTCTTCGCGCATCAGTTANNNCTTCTCAAA GATGAT	Site-directed random mutagenesis of *rhaR* K36 (middle downstream)
2982	GAACTCTTCTGCGCCACCGTGGCTACCTC GGCCAGAG	Site-directed random mutagenesis of *rhaR* L35, K36 (middle upstream)
2983	GCTCTCTTCTTAAAACTTNNNAAAGATGATTTTT TGCCAGC GAC	Site-directed random mutagenesis of r*haR* L38 (middle downstream)
2984	TCTCTCTTCTTAAAACTTCTCNNNGATGATTTTT TTGCCAGC GAC	Site-directed random mutagenesis of r*haR* K39 (middle downstream)
2985	TCTCTCTTCTTAAAACTTCTCAAANNNGATTTTT TTGCCAGC GAC	Site-directed random mutagenesis of *rhaR* D40 (middle downstream)
2986	GAACTCTTCTTTAACTGATGCGCCACCGTGG	Site-directed random mutagenesis of r*haR* L38, K39, D40 (middle upstream)
2987	TCTCTCTTCCTCAAAGATNNNTTTTTTGCCAGC GACCAGCAG	Site-directed random mutagenesis of *rhaR* D41

		(middle downstream)
2988	TCTCTCTTCCTCAAAGATGATNNNTTTGCCAGC GACCAGCAG	Site-directed random mutagenesis of *rhaR* F42 (middle downstream)
2989	TCTCTCTTCCTCAAAGATGATTTTNNNGCCAGC GACCAGCAG	Site-directed random mutagenesis of *rhaR* F43 (middle downstream)
2990	GAACTCTTCTTGAGAAGTTTTAACTGATGCGC	Site-directed random mutagenesis of *rhaR* D41, F42, F43 (middle upstream)
2991	TCTCTCTTCGATTTTTTTNNNAGCGACCAGCAG GCAGTC	Site-directed random mutagenesis of *rhaR* A44 (middle downstream)
2992	TCTCTCTTCGATTTTTTTGCCNNNGACCAGCAG GCAGTCGCT	Site-directed random mutagenesis of r*haR* S45 (middle downstream)
2993	TCTCTCTTCGATTTTTTTGCCAGCNNNCAGCAG GCAGTCGCT	Site-directed random mutagenesis of *rhaR* D46 (middle downstream)
2994	GAACTCTTCAAATCATCTTTGAGAAGTTTTAAC TGATGCGCC	Site-directed random mutagenesis of *rhaR* A44, S45, D46 (middle upstream)
3010	GCGGAATTCCTTTTGTAGGTATAAGATGGAC TTTAAATACACTGAAGAAAAAGAAACAATA AAANNNAATAATATTATGATTCATAAA-3’	Upstream oligo to randomize *rns* I14
3011	GCGGAATTCCTTTTGTAGGTATAAGATGGAC TTTAAATACACTGAAGAAAAAGAAACAATA AAAATTNNNAATATTATGATTCATAAA-3’	Upstream oligo to randomize *rns* N15
3012	GCGGAATTCCTTTTGTAGGTATAAGATGGAC TTTAAATACACTGAAGAAAAAGAAACAATA AAAATTAATNNNATTATGATTCATAAA-3’	Upstream oligo to randomize *rns* N16
3015	CAAGAATTCATTTTGTAGGTATCCAATGTCATC TGTGAGTTTTTTTTCTG	Upstream oligo to amplify *rns* (157-265)
3016	ATGGACTTTAAATACACTGAAGAA	Upstream *rns* oligo with out Shine-Dalgarno seq. Oligo

		started with initiation site ATG
3060	GCACATATGGACTTTAAATACACTGAAG	Upstream oligo to amplify and clone *rns*, *rns* NTD in to pDZ1 and pDZ3
3061	ACGGGATCCTTATCCACCTTTAAAATAAGTG	Downstream oligo to amplify and clone *rns*, *rns* DBD in to pDZ1
3062	ACGGGATCCTTATATATAAATTGATTCTATTATTTTTTC	Downstream oligo to amplify and clone *rns* NTD to clone rns NTD in to pDZ1
3063	GCACATATGTCATCTGTGAGTTTTTTTTCTG	Upstream oligo to amplify and clone *rns* DBD in to pDZ1 and pDZ3
3064	ACGGGATCCTCCACCTTTAAAATAAGTGAAAAATTG	Downstream oligo to amplify and clone *rns*, *rns* DBD in to pDZ3
3065	ACGGGATCCTATATAAATTGATTCTATTATTTTTTC	Downstream oligo to amplify and clone *rns* NTD in to pDZ3
3075	AGCGGATAACAATTTCACACAGGA	Upstream oligo for sequencing pHG165 with insert
3076	CGCCAGGGTTTTCCCAGTCACGA	Downstream *lacZ* oligo for sequencing pHG165 with insert
3119	CATGATAACTCAGCTTTTATATCTAGCTTG	*rns* middle forward primer for sequencing
3120	GCTGTTTATATTTTTTAATTCATCCAACAAGG	*rns* middle reverse primer for sequencing
3121	GGTCGTCAGACTGTCGATGAAG	Upstream oligo for sequencing insert cloned into pMAL-c2X
3159	TTGCATTGTC**TTGTTTTTTTATC**TCATTAATC	IR-labeled top strand of *cfaAp* with Rns binding site (bold). For EMSAs.
3160	GATTAATGA**GATAAAAAAACAA**GACAATGCAA	IR-labeled bottom strand of *cfaAp* with Rns binding site (bold). For EMSAs.
3161	CACAATTCAGCAAATTGTGAACATCATCACATT	IR-labeled top strand with full

	CATCTTTCCCTG	CRP binding site for EMSA
3162	CAGGGAAAGATGAATGTGATGATGTTCACAATTTGCTGAATTGTG	IR-labeled bottom strand with full CRP binding site for EMSA
3207	GCGGAATTCCTTTTGTAGGTATAAGATGGACTTTAAATACACTGAAGAAAAAGAAACAATANNNATTAATAATATTATGATTCATAAA	Upstream oligo to randomize *rns* K13
3208	GCGGAATTCCTTTTGTAGGTATAAGATGGACTTTAAATACACTGAAGAAAAAGAAACANNNAAAATTAATAATATTATGATTCATAAA	Upstream oligo to randomize *rns* I12
3209	CAGGCTCTTCACTTGAAAATTCATACCAG	Middle upstream oligo to randomize *rns* residues 216, 217, 218, 219
3212	CAGGCTCTTCCAAGTAGTAATAACGCAGCNNNACTCATTCTCAA	Middle downstream oligo to randomize *rns* residues 216
3215	CGAGCTCTTCCAAGTAGTAANNNCGCAGCCTTACTCATTCTC	Middle downstream oligo to randomize *rns* residues 219
3221	CGCGGATCCTTTGTTTGCATTGTCTTGTTTTTTTATCTCATTT	Upstream oligo amplify *cfaAp* -60 to +360 relative to TSS
3222	CGCGAATTCCGCCTCAAAATATACTC	Downstream oligo amplify *cfaAp* -60 to +360 relative to TSS
3223	GCGGGATCCATAAGTGTAGTTTGCTTT	Upstream oligo amplify *cfaAp* -162 to +14 relative to TSS
3224	GCGCTCTTCTTTTATTAATGATTACCTGACT	Downstream oligo amplify *cfaAp* -162 to +14 relative to TSS
3227	GGAATCAATGCCTGAGTG	To identify single copy integrants pSE298/pSE299 into MC4100 at *att* $_{HK022}$ site
3228	ACTTAACGGCTGACATGG	To identify single copy integrants pSE298/pSE299 into MC4100 at *att* $_{HK022}$ site
3229	ACGAGTATCGAGATGGCA	To identify single copy integrants pSE298/pSE299 into MC4100 at *att* $_{HK022}$ site

3230	GGCATCAACAGCACATTC	To identify single copy integrants pSE298/pSE299 into MC4100 at *att* $_{HK022}$ site
3243	GCGGAATTCCTTTTGTAGGTATAAGATGGACTT TAAATACACTGAAGAAAAAG	Upstream *rns* oligo to randomize K249, H250, Y251, G252 and Y261
3244	GCGAAGCTTTTATCCACCTTTAAAATAAGT GAAAAATTGCTTTGGTGTAACACCATAATG NNNATTAAAAATCCT	Downstream oligo to randomize *rns* K249
3245	GCGAAGCTTTTATCCACCTTTAAAATAAGT GAAAAATTGCTTTGGTGTAACACCATANNNTTT ATTAAAAATCC	Downstream oligo to randomize *rns* H250
3246	GCGAAGCTTTTATCCACCTTTAAAATAAGT GAAAAATTGCTTTGGTGTAACACCNNNATG TTTATTAAAAATCC	Downstream oligo to randomize *rns* Y251
3247	GCGAAGCTTTTATCCACCTTTAAAATAAGT GAAAAATTGCTTTGGTGTAACNNNATA ATGTTTATTAAAAATCC	Downstream oligo to randomize *rns* G252
3248	GCGAAGCTTTTATCCACCTTTAAANNNAGT GAAAAATTGCTTTGG	Downstream oligo to randomize *rns* Y261
3326	TTGCATTGTCT**TGTTTTTTTATC**TCATTAATC	Non IR-labeled top strand of *cfaAp* with Rns binding site (bold). For EMSAs.
3331	GCAGAATTCATGGACTTCAAATACACGGAA GAAAAAG	Upstream oligo to amplify and clone *rns*^{co} *or rns*^{co} NTD into pMAL-c2X vector
3332	GCAGAATTCAGCAGCGTGAGCTTTTTCAGCG	Upstream oligo to amplify and clone *rns*^{co} DBD into pMAL-c2X vector
3333	CGACAGAAGCTTTTAAATGTAGATGCTTTCGAT G	Downstream oligo to amplify and clone *rns*^{co} NTD into pMAL-c2X vector
3334	CGACAGAAGCTTTTAGCCGCCTTTGAAGTAGGT G	Downstream oligo to amplify and clone *rns*^{co} or *rns*^{co} DBD into pMAL-c2X vector

3347	GGAAGGGAGATTGATGGTAG	*virB* amplification - qRT-PCR (forward primer)
3348	GAACTTCAAGATCTGCTCCTGC	*virB* amplification - qRT-PCR (reverse primer)
3349	CTCAATTCAACACTCTTTCACAG	*icsB* amplification - qRT-PCR (forward primer)
3350	GGCTGTACCGATGCCATGAAAACG	*icsB* amplification - qRT-PCR (reverse primer)
3353	GCGCTCTTCTAAAAGTGTTTTTTTATCAAA CTGACAACACATCATCTACGGACAGGGGCC GCATTATTGCTGGCCGGAATTCTGC	To construct *nlpA*^{cfaA}-*lacZ*. Had *nlpA* promoter +15 to +82 relative to *nlpA* TSS. Replaces *nlpA* Rns binding site with *cfaA* Rns binding site (upstream oligo)
3354	GCAGAATTCCGGCCAGCAATAATGCGGCCCCTG TCCGTAGATGATGTGTTGTCAGTTTGATAA AAAAACACTTTTAGAAGAGCGC	To construct *nlpA*^{cfaA}-*lacZ*. Had *nlpA* promoter +15 to +82 relative to *nlpA* TSS. Replaces *nlpA* Rns binding site with *cfaA* Rns binding site (downstream oligo)
3385	CTTTCGGGTACTCAAGAAC	*icsA* amplification - qRT-PCR (forward primer)
3386	GAGAAAGTCCATCAACAGG	*icsA* amplification - qRT-PCR (reverse primer)
3387	CTGCATTTTCAAACACAGC	*ipaB* amplification - qRT-PCR (forward primer)
3388	GAGTAACACTGGCAAGTC	*ipaB* amplification - qRT-PCR (reverse primer)
3393	AACGTCAATGAGCAAAGGTATTAA	*rrsA* amplification - qRT-PCR (forward primer)
3394	TACGGGAGGCAGCAGTGG	*rrsA* amplification - qRT-PCR (reverse primer)
3395	GCGGAATTCCTTTTGTAGGTATAAGATGGAC TTTAAATACACTGAAGAAAAAGAAACAATA AAAATTAATAATNNNATGATTCATAAATACAC	Upstream oligo to randomize *rns* I17

3396	GCGGAATTCCTTTTGTAGGTATAAGATGGAC TTTAAATACACTGAAGAAAAAGAAACAATA AAAATTAATAATATTNNNATTCATAAATACAC	Upstream oligo to randomize *rns* M18
3398	CAGGCTCTTCTATTACTACTTGAAAATTCATAC	Middle upstream oligo to randomize *rns* residues 211, 215
3399	CAGGCTCTTCTAATAACGCAGCCTTNNNCATTC TCAATTGCAT	Middle downstream oligo to randomize *rns* residue 215
3400	CAGGCTCTTCTAATAACGCAGCCTTACT CATTCTCAANNNCATTAATATCTGATT	Middle downstream oligo to randomize *rns* residue 211
3403	GCGTCGCCGCTTTCATCG	Downstream oligo to sequence inserts cloned into pHKLac1 vector
3411	GCAGAAGGCCATCCTGAC	Upstream oligo to sequence inserts cloned into pHKLac1 vector

Undelined sequences are primer-template mismatches that include restriction endonuclease cleavage sites and random nucleotides.

Construction of a *cfaA-lacZ*$^{\Delta 60}$ fusion with single Rns binding site. *cfaA-lacZ*$^{\Delta 60}$ fusion has *cfaA* promoter region from -60 to +360 [relative to *cfaA* transcription start site (TSS)] fused to a *lacZ* reporter gene. Strain SME4324 carries a single copy of this fusion integrated into the chromosome at *att* $_{HK022}$ site. To construct this strain, *cfaA-lacZ* promoter region from -60 to +360 was amplified from strain SME3749 using forward primer 3221, 5'-CGCGGATCCTTTGTTTGCATT GTCTTGTTTTTTTATCTCATTT -3' and reverse primer 3222, 5'- CGCGAATTCCGCCTCAAAATATACTC -3' (Underlined regions in the primers are the primer-template mismatches and includes sites for restriction endonucleases digestion), digested with *Eco*RI and *Bam*HI and then ligated into the same sites of pHKLac1 (39). The resulting plasmid pSE298 was confirmed by sequencing for the correct insertion of the *cfaA-lacZ* promoter region and then integrated into the chromosome of *E. coli* K-12 strain MC4100 (206) using a method described before (204). MC4100 was transformed with pSE298 and a temperature sensitive helper plasmid pAH69 (204) that helped in the integration of pSE298 into the chromosome of MC4100. Plasmid pSE298 had a phage attachment site att_{HK022} that helped in attachment of plasmid at bacterial attachment site (*attB*) in the MC4100 chromosome. Plasmid pAH69 was a temperature sensitive plasmid and had an *int* gene that encoded an integrase enzyme. Integrase enzyme helped in the integration of pSE298 into the chromosome of MC4100. The transformation mix was suspended in TY, incubated at 37°C for 60 min (for the expression of integrase) and 42°C for 30 min (for curing helper plasmid), spread on a spectinomycin plate and incubated at 37°C for 24 h. Colonies were restreaked twice on nutrient agar plates and once on selective spectinomycin plate at 42°C to select for the colonies that lost the helper plasmid and have a stable pSE298 integrant into MC4100 chromosome. Colony PCR was then performed to identify single copy plasmid integrants.

Construction of an *nlpA*^{cfaA}*-lacZ* fusion. The *nlpA*^{cfaA}*-lacZ* reporter fusion is a transcriptional fusion of the *nlpA* promoter region from -162 to +82 (relative to the *nlpA* TSS) fused to the *lacZ* reporter gene. In addition, the Rns binding site that is native to the *nlpA* promoter region with the sequence of the single Rns binding site in the *cfaA-lacZ*$^{\Delta 60}$ fusion. To construct strain SME4474, which carries a single copy of the *nlpA*^{cfaA}*-lacZ* fusion, two sets of primers were used. One set of primers (forward primer 5'- GCGGGATCCATAAGTGTAGTTTGCTTT -3' and reverse primer 5'- GCGCTCTTCTTTTATTAATGATTACCTGACT -3') amplified the *nlpA* promoter region between -162 to +14 from strain AP76. Another set of (forward primer 5'- GCGCTCTTCTAAAAG**TGTTTTTTTAT**CAAACTGACAACACATCATCTACGGACAGG GGCCGCATTATTGCTGGCCGGAATTCTGC -3' and reverse primer 5'- GCAGAATTCCGGCCAGCAATAATGCGGCCCCTGTCCGTAGATGAT

GTGTTGTCAGTTTG**ATAAAAAAACAC**TTTTAGAAGAGCGC -3') were hybridized (without PCR amplification) to complete the remainder of the modified *nlpA* promoter region. These oligos replaced the Rns binding site sequence at *nlpA* (+15 to +26, indicated by bold letters) with the specified Rns site at *cfaA*, and carried the native *nlpA* sequence from +27 to +82. The DNA fragment carrying the -162 to +14 region was digested with *Bam*HI and *Ear*I, and the DNA fragment carrying the +15 to +82 region was digested with *Ear*I and *Eco*RI. The *Ear*I restriction endonuclease sites allowed the two DNA fragments to be joined seamlessly (with no additions or deletions in the gene) (81). The two DNA fragments were ligated into plasmid pHKLac1 (39) digested with *Bam*HI and *Eco*RI to create the transcriptional fusion with *lacZ* in plasmid pSE299. The cloned region was sequenced to confirm the correct insertion of the *nlpA*^{cfaA} promoter region and the plasmid pSE299 was integrated into the chromosome of MC4100 (206) using a method identical to the integration of pSE298 into the chromosome of

MC4100 (described in the previous page). Single copy plasmid integrants of pSE299 into the chromosome of MC4100 were identified by colony PCR to make strain SME4474.

Site-directed random mutagenesis of *rns* RS2 and AS2 regions. Site-directed random mutagenesis of the Rns RS2 and AS2 regions was performed by PCR using primers that incorporated NNN at one codon at a time of each of these regions. To mutagenize codons that were near the 5' (encoding RS2 positions 12-18) or 3' ends (encoding AS2 positions 249-252, 261) of the *rns* gene, two primers were used to amplify the *rns* gene, with either the forward or reverse primer containing three random nucleotides (NNN) at the position to be mutagenized, and the other primer containing wild-type sequence. For mutagenesis of positions that were not near the ends of the *rns* gene (encoding AS2 positions 211, 215, 216 and 219), four primers were used, with pairs of oligos amplifying the *rns* gene in two fragments, and the middle downstream primer containing random nucleotides (NNN) at the position to be mutagenized. The amplified DNA fragments were digested with *Ear*I, *Eco*RI and *Hind*III and cloned into the *Eco*RI and *Hind*III sites of the moderate copy number plasmid pHG165 (205), by which *rns* was expressed from a *lac* promoter located in the pHG165 (205). The resulting plasmids were individually transformed into SME4324 and plated onto nutrient agar plates with X-gal and ampicillin. Variants with a range of activity levels relative to wild-type Rns were picked, restreaked to get pure cultures, and assayed for transcription activation of *cfaA-lacZ*$^{\Delta 60}$ by β-galactosidase assay. Variants were sequenced on both the strands to identify the mutation and to ensure that there were no additional mutations.

β-galactosidase assays in 96-well plates. Starter cultures of strains to be assayed were grown overnight in tryptone-yeast extract broth (TY) supplemented with ampicillin. 10 µl of each overnight grown culture was inoculated into 2 ml of TY medium supplemented with ampicillin,

grown at 37^{o}C with shaking in 13 X 100 mm tubes to an OD_{600} of ~0.4, and placed on ice. A 10 µl sample of each cell culture was added to each of two wells of a 96-well microtitre plate and diluted 10-fold by adding 90 µl of TY medium. A 50 µL aliquot of lysis/ONPG (*o*-nitrophenyl-β-D-galactopyranoside) buffer [3 parts of ZOB buffer (11) to 1 part of 10 mg lysozyme (Sigma, St. Louis, MO) dissolved in PopCulture cell lysis reagent (EMD, Gibbstown, NJ)] was added to each well and incubated at room temperature. Reactions were stopped by adding 50 µL of 1M Na_2CO_3 (208). OD_{420} and OD_{550} values were measured in a PowerWave XS plate reader equipped with KC^4 data analysis software (BioTek Instruments). The OD_{600} values for each cell culture were obtained from 200 µl of each culture in another 96-well plate. The OD_{420}, OD_{550} and OD_{600} readings were corrected by subtracting the readings at the same wavelength obtained for control samples that had no bacterial cells added. β-galactosidase activities were measured by in Miller units using the formula below (208).

$$\text{Miller units} = (OD_{420} - OD_{550})*1000/(OD_{600}*t*V) \qquad \text{Eqn (1)}$$

where t = time of reaction (in minutes), and V = volume of cell suspension added to the reaction (in milliliters)

Specific activities (Miller units) were averaged from at least three independent experiments with two replicates in each experiment. Activity was expressed as percentage of wild type activity. Error bars represent the standard deviation of the mean values.

Western blotting. Rabbit polyclonal anti-Rns antibodies were raised commercially (Cocalico Biologicals, Reamstown, PA), by immunizing rabbit five times with a total of 350 µg of the purified MBP-Rns protein. Anti-Rns antibodies from the serum were affinity purified using purified MBP-Rns and used as primary antibody. Cultures expressing wild-type Rns or variants

were assayed by collecting 1 ml from bacterial cultures that were grown to OD_{600} of ~0.4 for β-galactosidase assays. The cells were sedimented and resuspended in 100 µl of 1X SDS buffer (0.191M glycine, 0.0247 M tris base, 0.1% SDS), mixed with 25 µl of 5X loading dye (3.75% trizma base, 10% SDS, 25% (v/v) 2-mercaptoethanol, 50% (v/v) glycerol, 0.5% bromophenol blue), boiled for 10 min, and 20 µl was loaded onto 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gel. Samples were electrophoresed, transferred onto nitrocellulose membranes, blocked with 5% non-fat dry milk and incubated with primary and secondary antibodies using standard procedures. Antibodies used in the Western blot were as follows; Rabbit anti-Rns primary antibody (1:500 dilution), Mouse anti-DnaK primary antibody (1:10,000) (Invitrogen, Grand Island, NY), IR 680-labelled anti-rabbit secondary antibody (1:10,000) and IR 800-labelled anti-mouse secondary antibody (1:10,000) (Abcam, Cambridge, MA). The blots were imaged using an Odyssey Infrared Imaging System (LI-COR, Lincoln, NE) and analyzed using ImageJ software (209).

Rns structural modeling. A computational model for three-dimensional structure for Rns was generated using I-TASSER structural prediction program (157). The obtained structure was displayed in ribbon using PyMOL program (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

Construction of IPTG inducible Rns plasmids for protein overexpression and purification. Plasmids were constructed to express Rns, Rns-NTD and Rns-DBD with GB1 tag or maltose binding protein (MBP) tag. To construct plasmids with GB1 at the N-terminus of the protein pDZ1 plasmid was used. Rns and Rns-NTD were PCR amplified (Rns: primers 3060 and 3061; Rns-NTD: primers 3060 and 3062) and cloned into *Nde*I and *Bam*H1 sites of pDZ1 plasmid. To construct plasmids with GB1 at the C-terminus of the protein, pDZ3 plasmid was used. Rns,

Rns-NTD and Rns-DBD were PCR amplified with primers (Rns: primers 3060 and, 3064; Rns-NTD: primers 3060 and, 3065; Rns-DBD: primers 3063 and, 3064) and cloned into *Nde*I and *Bam*H1 restriction sites of pDZ3 plasmid. Plasmids with codon-optimized version of *rns* (*rns*^{co}) with GB1^{basic} at either N-terminus or C-terminus of the protein were also generated. To make N-terminal GB1^{basic} version of *rns*^{co}, *rns*^{co} was PCR amplified with primers 3264 and 3265 using pSU18/*rns* as template and cloned into pSE290 plasmid at *Nde*I and *Bam*H1 sites. To make C-terminal GB1^{basic} version, *rns*^{co} was PCR amplified with specific primers (3264, 3266) using pSU18/*rns* as template and cloned into pSE295 plasmid at *Nde*I and *Bam*H1 sites.

Plasmids that express IPTG inducible N-terminal MBP tagged versions of Rns^{co}, Rns-NTD^{co} and Rns-DBD^{co} were constructed by amplification of insert using primers (Rns^{co}: primers 3331 and, 3334; Rns-NTD^{co}: primers 3331 and, 3333; Rns-DBD^{co}: primers 3332 and, 3334) and cloning downstream and in-frame with *malE* (encoding maltose binding protein, MBP) in the vector pMAL-c2X (New England Biolabs, Beverly, MA) at *Eco*RI and *Hind*III restriction sites. Plasmids that express IPTG inducible Rns and VirF expressed from pHG165 plasmid were constructed by Jeff Skredenske. Rns and VirF were amplified and cloned into pHG165 at *Eco*RI and *Hind*III restriction sites.

Rns-DBD-GB1-His6 purification. *E. coli* strain SME3863 was used for the overexpression and purification of Rns-DBD-GB1-His6 using Ni^{2+} affinity chromatography column at 4^{o}C. In brief, a starter culture of 5 ml TY supplemented with ampicillin and kanamycin was inoculated with SME3863 and grown overnight at 37^{o}C. A 1:100 dilution of bacteria was further inoculated into fresh medium (usually 100-200 ml), allowed to grow until OD_{600} of ~0.8, induced with 1 mM of isopropyl- β-D-1-thiogalactopyranoside (IPTG) at 15^{o}C and incubated ~18 h with shaking. After overnight induction cells were harvested by centrifugation at 4°C and concentrated to > 40-fold

in ice-cold binding buffer [20 mM Tris-base, 500 mM NaCl, 1 mM Imidazole, 0.1 mM tris (2-carboxyethyl) phosphine (TCEP), pH 7.9). Cells were added with 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.53 mg/ml Lysozyme and kept on ice for 15 min. Cells were lysed by either French press for 3 times or sonication (sonicator was programmed to operate 15 sec on and 15 sec off for a total of 4 min at 30% pressure). Cell lysates were subjected to DNase treatment (2.5 µg/ml DNase and 30mM $MgCl_2$) on ice for 15min and centrifuged (15,600 X g, 30 min) to remove cell debris. Rns-DBD-GB1-His6 was ~50% soluble and this soluble supernatant fraction was loaded to HiTrap immobilized Ni^{2+} affinity chromatography column (Sigma) connected to either BioLogic low pressure liquid chromatography system (Bio-Rad Laboratories) or AKTA flow pressure liquid chromatography (FPLC) system (GE Healthcare). The Ni^{2+} affinity chromatography column has a column volume of 5 ml and was prepared by charging the column with 5 column volumes (CVs) of 50 mM nickel sulphate and equilibration with 5 CVs of binding buffer. After running the sample, column was washed with 5 CVs of wash buffer (20 mM Tris-base, 500 mM NaCl, 60 mM Imidazole, 0.1 mM TCEP, pH 7.9) and eluted using 5 CVs of elution buffer (20 mM Tris-base, 500 mM NaCl, 500 mM Imidazole, 0.1 mM TCEP, pH 7.9).

In addition, protein was purified from pellet by using urea. This procedure was used in my initial purification experiments induction of cells at 37^{o}C yielded no soluble protein. The pellet fraction of sonicated lysates were resuspended in binding buffer having either 2 or 4 or 6 M urea and rocked overnight at 4^{o}C. I found that the protein was relatively more soluble in binding buffer with 6 M urea. The cell supernatant (6 M urea containing fraction) was loaded into Ni^{2+} affinity chromatography column and eluted in 6 M urea containing elution buffer (20 mM Tris-base, 500 mM NaCl, 500 mM Imidazole, 0.1 mM TCEP, pH 7.9).

Electrophoretic mobility shift assays with purified Rns-DBD-GB1-His6. Binding of purified Rns-DBD-GB1-His6 to a DNA fragment containing the *rns* binding site at *cfaA* promoter was analyzed by electrophoretic mobility shift assay (EMSA). DNA probes were generated by hybridizing the 5' IR-labeled 3159 and 3160 oligonucleotides (oligos, Eurofins MWG Operon, Huntsville, AL). For Hybridization, oligos were mixed in 1:1 ratio (usually 10 µl of 10 µM each oligo) in STE buffer (50 mM NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 8.0), incubated at 94°C for 2 min and gradually cooled to room temperature. EMSA gels (6% acrylamide/0.1% methyl-bis-acrylamide, 0.25x TBE) were made and pre-electrophoresed for at least 30 min (150 V) in 0.25x TBE buffer (22 mM Tris, 22 MM boric acid and 0.5 mM disodium EDTA) before loading the reactions. The temperature of the gel apparatus was maintained below 10°C by keeping dry ice in the gel box (Bio-Rad Laboratories). Reactions containing 1 µl of 1:15 dilution of labeled DNA (~0.11 µM final concentration) in STE buffer, 1.2 µl of 5x EMSA buffer [50 mM Tris-HCl (pH 7.4), 250 mM KCl, 5 mM dithiothreitol, 32.5% (v/v) glycerol, 0.5 mg/ml bovine serum albumin (BSA) and 50 µg/ml salmon sperm DNA] and 2.8 µl of water were incubated at 37°C for 5 min. Three different dilutions of 1 µl of Rns-DBD-GB1-His6 (9 µM, 3 µM, 1 µM final concentrations) were then added to reaction mixture and further incubated at 37°C for 10 min before loading onto the gel. DNA loading dye was added only to the free DNA sample. After electrophoresis (~45 min at 150 V) gels were scanned using an Odyssey infrared imager (LI-COR, Lincoln, NE).

MBP-Rns purification. *E. coli* strain SME4147 (KS1000/pRARE2/pMBP-Rns1) was used for the overexpression a fusion of the maltose binding protein to Rns (MBP-Rns). The fusion protein was purified using amylose affinity chromatography at 4^{o}C. In brief, SME4147 was grown aerobically at 37°C in TY medium supplemented with 200 µg/ml ampicillin, 30 µg/ml

chloramphenicol and 0.2% (w/v) glucose until an OD_{600} of ~0.4. A 1:100 dilution of bacteria was further inoculated into fresh medium, allowed to grow until OD_{600} of ~0.6, induced with 1mM of isopropyl- β-D-1-thiogalactopyranoside (IPTG) at 15^{o}C and incubated overnight with shaking. Cells were harvested by centrifugation at 4°C and concentrated 40-fold in ice-cold binding buffer [20 mM Tris-base, 500 mM NaCl, 1 mM EDTA, 0.1 mM tris (2-carboxyethyl) phosphine (TCEP) or 10 mM β-mercaptoethanol, pH 7.6). 1 mM phenylmethylsulfonyl fluoride (PMSF) and 0.53 mg/ml lysozyme were added to the cell suspension and it was lysed by freezing at -80°C for 15 min and thaw on ice and then sonicating twice. Cell lysates were subjected to DNase I treatment (2.5 µg/ml) on ice for 15min and centrifuged (15,600 x g, 30 min) to remove cell debris. The soluble supernatant fraction was loaded with a BioLogic LP chromatography system (Bio-Rad Laboratories) onto an amylose resin column (column volume 7 ml) (New England Biolabs) that had been pre-equilibrated with 80 ml of binding buffer. The column was washed with 120 ml of wash buffer (binding buffer plus 15% (v/v) glycerol) and eluted using elution buffer (wash buffer plus 10 mM maltose).

Gel filtration chromatography of purified MBP-Rns. The protein was further purified by gel filtration chromatography using a pre-packed analytical sephacryl 16/60 S-200 column (GE Healthcare). Gel filtration usually involved pre-equilibration of column with two CVs of either elution buffer [20 mM Tris-base, 500 mM NaCl, 1 mM EDTA, 0.1 mM TCEP, 15% Glycerol, 10 mM Maltose pH 7.6] or tris-buffer with KCl as salt and without maltose (15 mM Tris-base, 75 mM KCl, 5 mM β-mercaptoethanol, 10% Glycerol, pH 7.6), loading of up to 5 ml protein sample and collection of 2 ml fractions of peaks using "S-200 frac peak only" program. Molecular weight standards ranging from 10 kDa to 200 kDa proteins were passed through the column and used to determine the molecular weight of the eluted MBP-Rns protein.

Electrophoretic mobility shift assays with purified MBP-Rns. EMSAs were performed for the MBP-Rns to test whether the purified protein is active and binds to its specific labeled DNA. EMSA reactions were prepared as similar to EMSA reactions for Rns-DBD-GB1-His6 protein with the following modifications. Instead of using both IR-labeled *rns* specific oligonucleotides (3159 and 3160), a non-IR labeled top oligonucleotide (3326) and IR labeled bottom oligonucleotide (3160) were used. DNA and protein concentrations of ~2 nM and ~0.3 µM were used, respectively.

Affinity purification of Rns antibodies. Rabbit polyclonal anti-Rns sera were raised commercially (from Cocalico Biologicals, Reamstown, PA) by immunizing rabbit five times with a total of 350 µg of the purified MBP-Rns. Rns antibodies from the serum were affinity purified against purified MBP-Rns using the following protocol. 15-20 µg of the purified MBP-Rns protein was run on SDS-PAGE gel and electrophoretically transferred to nitrocellulose membrane using the standard procedure. Protein bands corresponding to the purified MBP-Rns was cut from the membrane after staining with ponceau S stain. Membrane strip was cut into small pieces and blocked at room temperature for 2 h in 15 ml blocking buffer [5% non-fat dry milk in 1x PBST (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4; pH of the mixture was adjusted to 7.4 and then added with 0.05% tween-20] using a 50 ml conical flask. Blocking buffer was then replaced with 20 ml of 1x PBST to which 1 ml of the rabbit serum sample was added. The resulting mix was rocked overnight at room temperature to facilitate the binding of Rns antibodies to the purified MBP-Rns on the membrane. Membrane strips were washed twice with 1x PBST and once with 1x PBS (15 min each) and rocked vigorously by adding 3 ml of 100 mM glycine at low pH (2.5) to facilitate the removal of Rns antibodies from the membrane. The pH of antibody solution was then adjusted to near neutral by adding 300 µl

of 1M Tris (pH 8.0). To ensure the correct detection of Rns by the affinity purified antibodies, Western blots were performed for *E. coli* strains SME4336 (expresses Rns from pHG165) and SME4337 (pHG165 control) using a standard Western blot procedure describe above. Rns antibodies were stored at -20^{o}C for long term storage.

Crystallization trials. The purified MBP-Rns protein was concentrated using a Millipore centricon (50 kDa cut off). MBP-Rns in the Tris elution buffer [20 mM Tris-base, 500 mM NaCl, 1 mM EDTA, 0.1 mM TCEP, 15% Glycerol, 10 mM Maltose pH 7.6] showed visible aggregation at higher concentrations (usually after 2 mg/ml). MBP-Rns in the Tris buffer with KCl as salt and without maltose (15 mM Tris-base, 75 mM KCl, 5 mM β-mercaptoethanol, 10% Glycerol, pH 7.6) did not show any aggregation even at concentrations as high as ~20mg/ml and thereby used for setting up crystal screens. Almost 400 different commercially available crystallization conditions (Wizard I, II, III, IV screens, Hampton I, II, PEG/Ion and salt RX screens) covering wide range of buffers and pH were tested.

Effect of SE-1 on DNA binding by purified MBP-Rns protein. Binding of purified MBP-Rns to a DNA fragment containing the Rns binding site at *cfaA* promoter region was analyzed by electrophoretic mobility shift assay (EMSA), in the absence or presence of SE-1. SE-1 (1-butyl-4-nitromethyl-3-quinolin-2-yl-4H-quinoline, formerly OSSL_051168) was obtained from eMolecules, Inc. (Solana Beach, CA; catalog #3761-0013) or Princeton BioMolecular Research (Princeton, NJ; catalog # OSSL_051168) and was dissolved in 100% DMSO. DNA probes were generated by hybridizing the oligonucleotides 3326 (non-IR labeled top oligonucleotide) and 3160 (IR labeled bottom oligonucleotide). Reactions containing 0.3 µl (~2 nM) of DNA, 0.6 µl of either 100% DMSO or various concentrations of OSSL_051168 (10.1, 20.3, 40.6, 81.25, 162.5, 325 and 650 µM) and 1.2 µl 5x EMSA buffer [50 mM Tris-HCl (pH 7.4), 250 mM KCl, 5

mM Dithiothreitol, 32.5% (v/v) glycerol, 0.5 mg/ml bovine serum albumin (BSA) and 50 μg/ml salmon sperm DNA] and 3.9 μl water were incubated at 37°C for 5 min. MBP-Rns (1 μl, ~0.3 μM) was added to the reaction mixture and further incubated at 37°C for 10 min before loading onto the gel. DNA loading dye was added only to the free DNA sample. Electrophoresis was performed in 0.25x TBE (final concentrations: 22.25 mM Tris base, 22.25 mM boric acid, 500 μM disodium EDTA, pH 8.3) using a non-denaturing 6% polyacrylamide gel (6% acrylamide, 0.1% *N*,*N*-methylenebisacrylamide) that was cooled to below 10^{o}C and pre-ran at ~150 V for 30 min. All EMSA reactions, including those without inhibitor, had a final concentration of 10% DMSO (the SE-1 solvent). Gels were scanned using an Odyssey infrared imager (LI-COR, Lincoln, NE), and quantified using Odyssey software, version 3.0.30. Graphs were drawn using Prism (GraphPad, La Jolla, CA).

***In vivo* dose-response experiments in *E. coli*.** Cultures of SME4382, SME4383 and SME3359 grown overnight were diluted to an OD_{600} of ~0.1 in MOPS-buffered minimal medium with ampicillin. Bacterial cultures were mixed with varying concentrations of SE-1 (2.2% DMSO, final), induced with 6.5 mM IPTG for 3 h at 37^{o}C. After 3 h of IPTG induction, cells were lysed with 50 μl ONPG/Lysis buffer (three parts ZOB (11) to one part 10 mg/mL lysozyme [Sigma, St. Louis, MO] dissolved in PopCulture cell lysis reagent [EMD Chemicals, Gibbstown, NJ]) and β-galactosidase activity was measured at OD_{420} in a PowerWave XS plate reader (BioTek Instruments, Inc., Winooski, VT). Uninduced (no IPTG and no SE-1) and uninhibited (6.5 mM IPTG and no SE-1) controls were included for each strain and used as a normalizer for β-galactosidase activity. IC_{50} values were calculated and graphs were drawn using Prism (GraphPad, La Jolla, CA). Error bars represent the standard error of the mean of three independent experiments with two replicates each.

Heterologous expression and purification of VirF. The IPTG inducible MBP-VirF expression plasmid, pMAL*virF*, was constructed by cloning *virF* downstream and in-frame with *malE* (encoding maltose binding protein, MBP) in the vector pMAL-c2X (New England Biolabs, Beverly, MA). MBP-VirF protein (referred to as VirF for simplicity) was purified using amylose affinity chromatography at 4^{o}C. Briefly, pMAL*virF* was transformed into *E. coli* strain KS1000 (New England Biolabs). The cells were grown to an OD_{600} of 0.5, transferred to 15^{o}C, 0.1 mM IPTG was added, and then incubated overnight with shaking. Cells were harvested and resuspended in 10 ml of binding buffer (20 mM Tris, 500 mM NaCl, 1mM EDTA, 1mM DTT, pH 7.4). Cells were lysed by three freeze thaw cycles with lysozyme (0.4 mg/ml), tris (2-carboxyethyl) phosphine (TCEP, 1 mM) and phenylmethylsulfonyl fluoride (PMSF, 1 mM), at -80^{o}C, followed by sonication and then centrifuged to remove cell debris. The supernatant was applied with a BioLogic LP chromatography system (Bio-Rad Laboratories) to an amylose resin column (New England Biolabs), pre-equilibrated with 80 ml binding buffer, the column was washed with 120 ml binding buffer. VirF was eluted with 40 ml of elution buffer [binding buffer plus 15% glycerol (w/v) and 10 mM maltose], and used directly for subsequent assays. The purified protein was greater than 90% pure, although some older preps showed partial breakdown to MBP and non-fusion VirF. EMSAs did not show any evidence of non-fusion VirF binding to DNA (data not shown), suggesting that the non-fusion VirF protein was inactive (likely due to aggregation).

Effect of SE-1 on DNA binding by purified MBP-VirF protein. Binding of purified VirF to a DNA fragment containing the VirF binding site at *virB* was analyzed by electrophoretic mobility shift assay (EMSA), in the absence or presence of SE-1. DNA probes were generated by hybridizing the following oligonucleotides: A short IR700-labeled LUEGO oligo (5'-

GTGCCCTGGTCTGG-3') (210), a top oligo containing the *virB* binding sites (underlined) (5'-AGAATATTATTCTTTTATCCAATAAAGATAAATTGCATCAATCCAGCTATTAAAATAGTA-3'), and a bottom oligo that was complementary to both of the preceding oligos (5'-TACTATTTTAATAGCTGGATTGATGCAATTTATCTTTATTGGATAAAAGAATAATATTCT CCAGACCAGGGCAC-3'). Reactions and electrophoresis were performed as described for MBP-Rns EMSAs with SE-1. Gels were scanned using an Odyssey infrared imager (LI-COR, Lincoln, NE), and quantified using Odyssey software, version 3.0.30. Graphs were drawn using Prism (GraphPad).

***Shigella virB-lacZ* reporter gene assays.** Colonies of *S. flexneri* strain BS536 that were red on TSA plates containing Congo red were grown overnight in Luria-Bertani broth at 30°C to maintain low basal expression from the *virB* promoter. The overnight grown cultures were diluted 1:100 into 10 ml of the same medium with either 0.3% DMSO or different concentrations of SE-1 dissolved in DMSO (final concentration 0.3% DMSO). Cultures were grown at 37°C in a shaking incubator to an OD_{600} of ~0.4, centrifuged to pellet the cells and then resuspended in Z buffer (61 mM Na_2HPO_4, 40 mM NaH_2PO_4-H_2O, 10 mM KCl, 1 mM $MgSO_4.7H_2O$) for a two-fold concentration of the cells (208). Control cultures were grown at 30°C, a temperature at which *virF* expression is not induced (37), to illustrate the basal level of expression from the *virB* promoter region. β-Galactosidase assays were performed as described below. Each cell culture with two different volumes of cell suspension (usually 20 µl and 40 µl) were added to two nonsterile 100 x 13 mm tubes and made to a final volume of 1 ml by adding working buffer. Cells were permeabilized by the addition of two drops of chloroform and one drop of SDS and subsequent vortexing for 5 sec. Cells were incubated at room temperature for 5-10 min, added with 200 µl of substrate ONPG (*o*-nitrophenyl-β-D-galactopyranoside) (4 mg/ml), and then

vortexed briefly to mix ONPG with the cell suspension. Reactions were incubated at room temperature until yellow color was clearly visible, but still light. Reactions were stopped by adding 500 µl of sodium carbonate (1M). 1 ml of each sample was centrifuged for 5 min, supernatant (~800 µl) was transferred to a disposable cuvette and OD_{420} was measured. OD_{600} was measured for each cell culture and β-galactosidase activity was measured in miller units.

Miller units = $(OD_{420}*1000)/(OD_{600}*t*V)$ Eqn (2)

t = time of reaction (in minutes), V = volume of cell suspension added to the reaction (in milliliters)

Three independent assays were performed with two replicates in each assay. Activities were presented as a percentage of the uninhibited, DMSO-only control sample. Error bars represent the standard error of the mean.

***Shigella* gene expression analysis.** Real-time quantitative reverse transcription PCR (qRT-PCR) was performed to test the effects of SE-1 on the expression of the VirF regulated genes *virB, icsA, icsB* and *ipaB*. The *gapA* and *rrsA* transcripts were used as internal controls. These are constitutively expressed *Shigella* genes commonly used to normalize mRNA levels (211-213). *S. flexneri* strain SME4331, which carries a null mutation in *ipgD* (*ipgD*$^-$), was grown at 30^{o}C overnight in TSB with ampicillin, diluted into the same medium to an OD_{600} of ~0.1, and 1ml aliquots further grown at 30^{o}C (DMSO only) or at 37^{o}C (SE-1 at 20 or 40 µM, or DMSO only, all 0.3% DMSO). Cells were grown to an OD_{600} of 1.0 (~2.5 h) and RNA was isolated using an RNeasy MiniElute Cleanup Kit (Qiagen, Germantown, MD), with DNA contamination eliminated using a Turbo DNase Kit (Ambion, Austin, TX). Complementary DNA (cDNA) was synthesized by random priming using a cDNA Reverse Transcription Kit (Applied Biosystems,

Foster City, CA). Primers (Table 2) were tested for PCR specificity using genomic DNA template to ensure a single amplicon of the expected size for each primer pair, and validated using 5-fold dilutions of cDNA (2000 to 0.0256 ng) to ensure a linear plot of cycle threshold (Ct) versus cDNA concentration. Validated primer sets were used to test mRNA profiles in reactions that contained 10 ng of cDNA, 10 µl of SYBR green master mix (Applied Biosystems), specific primers (0.5 µM each) and water to a final volume of 20 µl, using a Step One Plus Real-Time PCR System (Applied Biosystems). The data was analyzed using the $2^{-\Delta\Delta Ct}$ method and by Applied Biosystems real-time analysis software. The data analyzed by the $2^{-\Delta\Delta Ct}$ method is reported; however, both analyses produced comparable results. In brief, ΔΔCt values (see equation below) were separately obtained for each sample relative to two internal controls, *gapA* and *rrsA*.

$$[\Delta\Delta Ct = (Ct,_{Target} - Ct,_{Internal\ Control})_{SE\text{-}1} - (Ct,_{Target} - Ct,_{Internal\ Control})_{no\ SE\text{-}1}] \qquad \text{Eqn (3)}$$

The ΔΔCt values relative to the two internal controls for each sample were averaged and then used to calculate the relative expression levels ($2^{-\Delta\Delta Ct}$). As a control, we normalized each internal control to the other internal control. Control reactions used templates generated without reverse transcriptase to ensure low genomic DNA contamination. To ensure that products were not likely to be due to primer dimer artifacts, additional control samples were performed in which reaction mixtures contained primers alone (no template).

Testing the effect of SE-1 on growth of *E. coli* strains used for *in vivo* assays. To test the impact of SE-1 on growth of the bacterial strains used for *in vivo* assays (SME4382, SME4383 and SME3359), we compared growth rates in the presence and absence of SE-1. Briefly, the cells were grown at 37^{o}C with shaking for 8 h in MOPS-buffered minimal medium with glycerol and plus 6.5 mM IPTG, in 24-well plates, with or without: SE-1 (44 µM SE-1, 0.3% DMSO), or

DMSO (0.3%) in a PowerWave XS plate reader equipped with KC^4 data analysis software (BioTek Instruments). Readings were taken at 15 min intervals. Error bars represent the standard error of the mean.

Testing the effect of SE-1 on growth of *Shigella*. The *S. flexneri ipgD*$^-$ strain (SME4331) was grown overnight at 30°C and then diluted to an OD_{600} of ~0.1 in TSB supplemented with ampicillin. SE-1 was diluted in DMSO and added to the SME4331 cultures at final concentrations of 0, 20, or 40 μM SE-1 and 0.3% DMSO. These cultures were grown at 37°C in 24-well plates containing 1.2 ml of each sample in the temperature-controlled chamber of a PowerWave XS plate reader equipped with KC^4 data analysis software (BioTek Instruments). Bacterial growth was continued at 37°C for 8 h with shaking and OD_{600} was monitored every 5 min for growth curves (20 min time intervals were plotted).

Site-directed mutagenesis of RhaR Arm (residues 35-45) region. Wild-type *rhaR* and all mutants were cloned into and expressed from the plasmid pHG165 where expression was expressed from *lac* promoter on the vector. First, wild-type *rhaR* was cloned into pSU18 vector at *Eco*RI and *Hind*III restriction sites to make pSU18/*rhaR*, which was used as template for amplification of *rhaR* during mutagenesis. Site-specific random mutagenesis of *rhaR* was performed using PCR to make oligonucleotide-directed random mutations at each position in the proposed Arm region. Two sets of oligos were used to generate site-directed mutations at the desired position. One set of two oligos amplified a small fragment of *rhaR* and encoded random mutations (NNN) at the desired position. Another set of two oligos amplified the rest of the *rhaR* gene. Oligonucleotides that hybridized in the middle of the gene had *Ear*I restriction endonuclease recognition site and allowed seamless ligation of both the pieces of *rhaR* into pHG165. The ligation mix was transformed into the strain SME2525 [λϕ (*rhaS-lacZ*)Δ128

Δ(*rhaSR*)::Km *recA*::*cat*] and plated onto nutrient agar plates supplemented with X-gal, ampicillin and L-rhamnose. Variants with varied levels of activity relative to the wild type were picked, and restreaked to obtain pure colonies. Plasmid DNA was isolated from the variants, transformed into SME3160 strain [λϕ (*rhaS-lacZ*)Δ85 Δ(*rhaSR*)::Km *recA*::*cat*] and assayed for activation levels by β-galactosidase assay. Variants were sequenced on both strands to identify the mutation and to ensure that there were no additional mutations apart from the mutation at desired position.

β-galactosidase assays to measure activation levels of RhaR Arm variants. Starter cultures for β-galactosidase assays were grown by inoculating a single well isolated bacterial colony in 2 ml TY + ampicillin and incubating in a rotator for 7 h at 37^{o}C. 40 µl of this culture was further inoculated in 2.5 ml of pre-warmed overnight medium [MOPS buffered minimal medium with limiting glycerol (0.04%)] and grown overnight (15 h) in rotator at 37^{o}C. 100 µl of the overnight grown bacterial cultures were inoculated in 10 ml of pre-warmed growth medium (0.4% glycerol in MOPS buffered medium, with and without 0.4% sugar rhamnose) in 125 ml baffled flasks. Cultures were grown with vigorous shaking (~250rpm) until an OD_{600} of ~0.4 and stopped by keeping on ice. Cells were pelleted down by centrifugation, resuspended in working buffer (61 mM $NaHPO_4$, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 250 mM dithiothreitol), and assayed for β-galactosidase activity using the protocol outlined already (see above).

Specific activities were averaged for three independent experiments with two replicates in each experiment. Standard deviation of the mean was calculated for all the variants and used to determine the level of significance. β-galactosidase activities were expressed as percentage of wild type activity, with or without L-rhamnose.

Western blotting to quantify RhaR protein levels of variants. All the steps of Western blotting were carried as described above for Western blotting of Rns variants, except using RhaR antibodies in place of Rns antibodies. Antibodies used in the Western blotting were as follows; rabbit anti-RhaR primary antibody (1:500 dilution), mouse anti-DnaK primary antibody (1:10000) (Invitrogen, Grand Island, NY), IR 680-labelled anti-rabbit secondary antibody (1:10000) and IR 800-labelled anti-mouse secondary antibody (1:10000) (Abcam, Cambridge, MA). The blots were imaged using an Odyssey Infrared Imaging System (LI-COR, Lincoln, NE), quantified using ImageJ software and presented as percentage of wild-type RhaR protein levels.

Chapter 3

Mechanism of transcription activation in Rns: Role of residues in RS2 and AS2 regions in transcription activation

Variants in the Rns NTD RS2 region affect Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression.

Previous studies by Basturea *et al* showed that NTD of Rns is required for DNA binding and transcription activation (156). In addition, they also found that substitutions near the amino-terminus of Rns, I14T and N16D, alter the Rns DNA sequence specificity (156) suggesting that residues at the amino terminus of the Rns may be required for Rns activity. In order to characterize the function of the NTD in Rns, we first defined the two domains of Rns.

Earlier it was predicted that the Rns NTD is from residues 1-100 and DBD is from residues 131-265 with a possible linker between the residues 101-130 (156). However, after comparison of DBDs of other AraC family members including AraC, MarA and ToxT (which are ~100 amino acids long, rather than the ~30 amino acid domain previously predicted), we hypothesized that DBD may extend from residue 157 to 265. We first made an Rns DBD construct that included residues 157-265 and assayed the activation levels of Rns(157-265) in comparison to the wild type and found that DBD does not activate transcription (Fig. 5). Our electrophoretic mobility shift assays using purified Rns-DBD showed that the DBD (from residues 157-265) is functional and is capable of binding DNA, suggesting that the construct 157-265 can serve as DBD.

Then using ClustalW method (157) we aligned a group of Rns homologs (done by Dr. Susan Egan) and found a substantially higher sequence identity of 74% among residues around I14 and N16 (I12-M18) compared to the overall sequence identity of 26% among NTDs of these proteins (Fig. 6). We called this highly conserved region of the Rns NTD (residues I12-M18) RS2, for its predicted role as a regulatory site that contacts subdomain 2 of the DBD, at least in AraC family proteins that responds to effector molecules (79, 100, 101). Then, by using I-TASSER predicted structure of Rns (157) (Fig. 7) and the sequence alignment of Rns with ToxT

(Fig. 8) we found that Rns RS2 region residues aligns with ToxT β2 sheet residues (Y20-Y26) that are positioned to make inter-domain interactions with the ToxT DBD (79). With this information, we hypothesized that the Rns RS2 region may play a role in maintaining Rns in its activating state, and therefore, we mutagenized residues in the Rns RS2 region by site-directed random mutagenesis to understand the role RS2 region plays in overall Rns function.

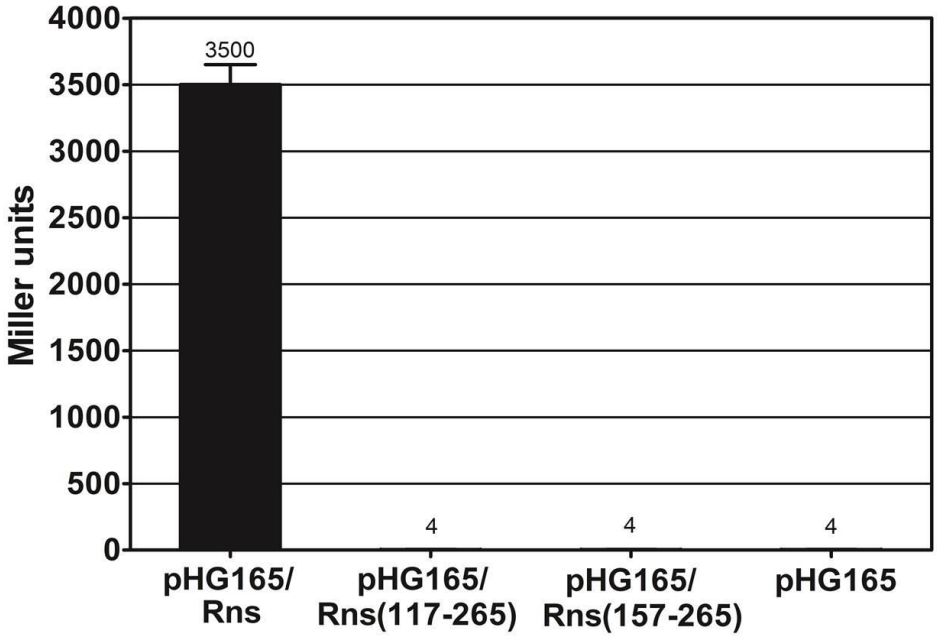


Fig. 5. N-terminal domain (NTD) of Rns is required for the Rns activation. Wild-type Rns and deletion constructs were expressed from a low copy number plasmid pHG165 and analyzed for their ability to activate transcription at *cfaA-lacZ* promoter region that has three Rns binding sites. Results shown were the average of three independent experiments with two replicates each. Error bars represent the standard deviation of the mean.

Fig. 6. Alignment of Rns RS2 region with the most closely related proteins. ClustalW method was used to align sequences. The over lined residues indicate RS region. Identical positions are red and marked "*"; strongly similar positions are green and marked ":"; weakly similar positions are blue and marked "."; different positions are black and unmarked. Gaps in the alignment are marked "-". Rns from ETEC, CfaD from ETEC 10407, AggR [*Escherichia coli* O104:H4], HdaR [*Escherichia coli*], CsvR [*Escherichia coli*], CS12 fimbria transcriptional activator protein [*Escherichia coli*].

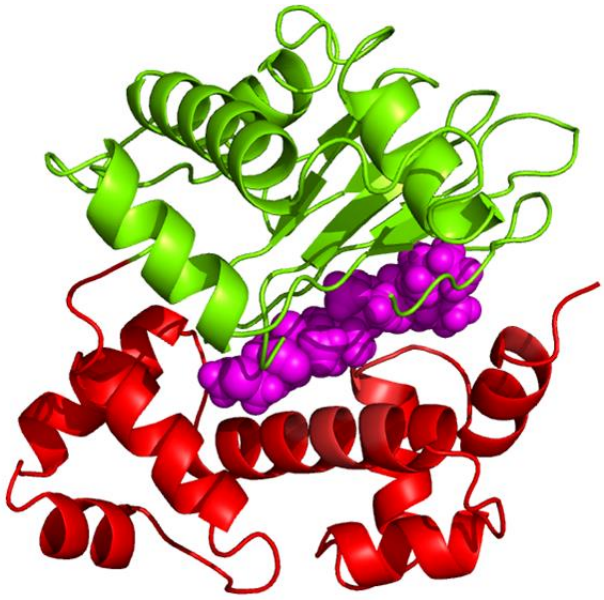


Fig. 7. Predicted Rns three-dimensional structure with RS2 region mapped. I-TASSER (157) was used to predict the 3D structure of the Rns. NTD was shown in green and DBD was shown in red. Residues corresponding to the RS2 region in the NTD were shown as purple spheres.

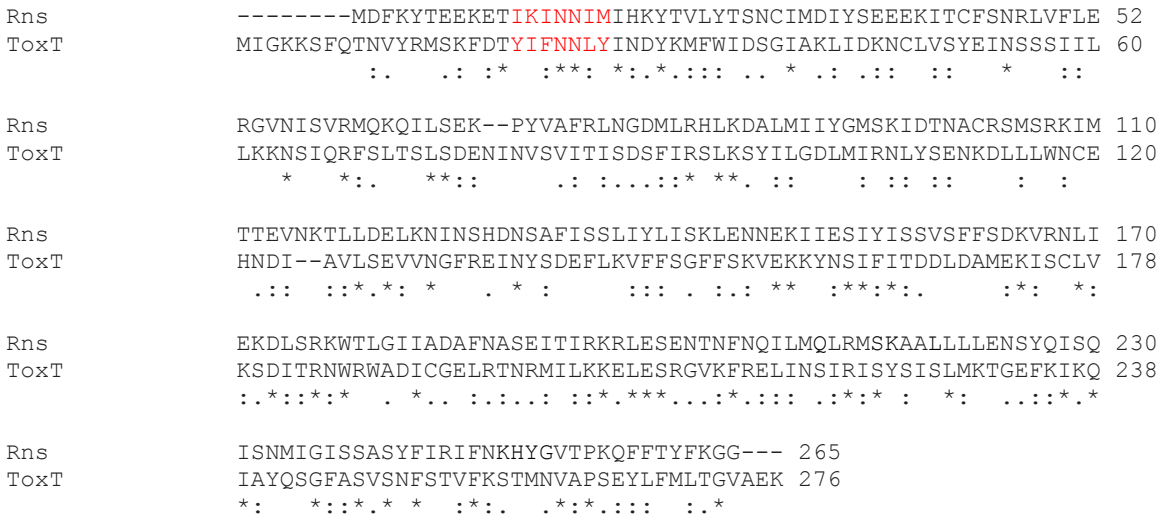


Fig. 8. Alignment of Rns withToxT. ClustalW method (157) was used to align sequences.

Residues in the RS2 region of Rns and the corresponding residues in ToxT were marked red.

We mutagenized the Rns RS2 region by randomizing one codon at a time (using oligonucleotides with NNN at the codon to be mutagenized) and generated an assortment of variants. Wild type and different variants of Rns (cloned into pHG165 vector) were tested for Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression levels at CFA/I promoter. The *cfaA-lacZ*$^{\Delta 60}$ fusion strain has a single Rns binding site that overlaps the -35 promoter element at which Rns is in a position to contact RNAP. The level of Rns dependent LacZ expression in this strain is consistently 90 fold higher for wild type compared to the strain transformed with vector only control (Table 4).

We isolated 25 unique variants at 7 positions in the Rns RS2 region. Out of the total 25 variants, 13 (52%) variants had significantly decreased Rns dependent LacZ expression. None of the variants increased Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression relative to wild type (Table 5). Variants that were defective for Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression were identified at every position in the RS2 region except K13 (Table 5). In principle, the decreases in the Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression could be due to decreased Rns activity or reduced Rns protein levels (resulting from poor protein expression or protein instability). To test this possibility, we performed Western blots for all the mutant strains, using antibodies raised against purified MBP-Rns protein. Figures 9 and 10 show the protein levels of different variants isolated from the RS2 region. Defective variants from I14, N15, N16 and I17 showed Rns protein levels similar to wild type (Table 5), suggesting that the defects observed in these variants were due to the protein activity and cannot be attributed to lower protein levels. However, all the defective variants isolated from I12 (I12R, I12G, I12E and I12W) (Fig. 9) and M18 (M18P and M18R) (Fig. 10) had significant decrease in protein levels (Table 5). The remainder of the substitutions at I12 (I12S) and M18 (M18G) had no significant defects in either LacZ expression or protein levels. Given that many of the variants had decreased protein levels at I12 and M18, these residues may

be required to stabilize Rns. However, the direct involvement of I12 and M18 in Rns dependent LacZ expression could not be concluded as all the defective variants had defects in protein levels (Table 5).

Despite having substitutions with various physico-chemical properties, none of the K13 variants showed any Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression defects, indicating that this residue was not important for Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression (Table 5). We could not make any definitive conclusions regarding the importance of I14 residue as we were able to isolate only one variant at this position, although the only variant isolated (I14W) was defective for Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression (Table 5).

Overall, we isolated multiple defective variants at positions N15, N16 and I17, suggesting that these residues are important for Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression (Table 5). Altogether, this data demonstrates that specific residues in the RS2 region of NTD are required for the transcription activation by Rns.

Distinguishing DNA binding defects from transcription activation defects in RS2 region. In principle, Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression defects observed in Rns RS2 region could be the result of either defects in DNA binding or direct defects in transcription activation (In this context, we refer to transcription activation as processes such as contacts with RNA polymerase, alteration of promoter DNA conformation and relief of repression). In order to further characterize the variants and identify the nature of these defects (DNA binding vs direct activation), assays were performed to test their ability to repress the transcription in a repression strain that carried *nlpA*^{cfaA}*-lacZ* fusion (Table 4). Repression of LacZ expression in *nlpA*^{cfaA}*-lacZ* fusion strain requires only binding of Rns to *nlpA* promoter region, unlike the LacZ expression

from *cfaA-lacZ*$^{\Delta 60}$ which requires both DNA binding and direct transcription activation to activate transcription. Therefore, comparison of results from both the fusion strains can differentiate variants that impact DNA binding versus direct transcription activation. Rns variants that are defective for both repression of LacZ expression at *nlpA*^{cfaA}*-lacZ* and activation of LacZ expression from *cfaA-lacZ*$^{\Delta 60}$ likely indicate defects in DNA binding. Rns variants with defects only in Rns dependent LacZ expression from *cfaA-lacZ*$^{\Delta 60}$ (wild-type repression levels *nlpA*^{cfaA}*-lacZ*) likely indicate direct defects in transcription activation. To eliminate the sequence-specific differences in Rns binding to the *nlpA* and *cfaA* promoter regions (due to differences in the Rns binding site sequences), we replaced the Rns binding site at *nlpA* promoter with one that is identical to the one found in *cfaA* promoter region in *nlpA*^{cfaA}*-lacZ* fusion. In *nlpA*^{cfaA}*-lacZ* fusion strain, wild-type Rns represses LacZ expression by ~2 fold compared to the strain transformed with vector only control (unrepressed state).

Our analysis of RS2 region variants for repression at *nlpA*^{cfaA}*-lacZ* and Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression levels and protein levels for the same variants have led to the following findings. All of the variants with wild type expression from the *cfaA-lacZ*$^{\Delta 60}$ showed wild-type protein levels and repression levels at *nlpA*^{cfaA}*-lacZ*, indicating no significant defects in those variants (Table 5). With the exception of I14W, all of the variants in the RS2 region that had defects in *cfaA-lacZ*$^{\Delta 60}$ expression showed protein levels and repression levels similar to the wild type. This suggests that these variants primarily impact activation function and the wild type residues at these positions (N15, N16 and I17) are required for full transcription activation by Rns (Table 5). The only substitution isolated at residue I14 (I14W) was defective for both *cfaA-lacZ*$^{\Delta 60}$ expression and *nlpA*^{cfaA}*-lacZ* repression indicating defects in DNA binding. However, the defects in *cfaA-lacZ*$^{\Delta 60}$ expression (~17 fold) for this substitution was much more severe than the

defects in *nlpA*^{cfaA}-*lacZ* repression (~2 fold). These results suggest that I14 may impact both DNA binding and transcription activation, although with only one variant, we cannot draw a strong conclusion (Table 5).

Overall, our analysis of variants in both *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}-*lacZ* strains suggested that three residues in RS2 region (N15, N16 and I17) were required for transcription activation function of Rns.

Table 4. Rns dependent activation of *cfaA-lacZ* and repression of *nlpA*^{cfaA}*-lacZ*.

lacZ reporter strain	Promoter[a]	Vector only control[b]	Rns[c]
cfaA-lacZ$^{\Delta 60}$	-60 to +360	40 ± 5	3650 ± 150
nlpA^{cfaA}*-lacZ*	-160 to +82	60 ± 1	30 ± 2

[a]Represents promoter sequence relative to the transcription start site

[b]β-galactosidase activity (Miller units) from fusion strains transformed with vector only control (pHG165)

[b]β-galactosidase activity (Miller units) from fusion strains transformed with pHG165 plasmid cloned with *rns*.

Miller unit values were presented as mean ± standard deviation of three samples.

Table 5. Transcription activation of *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ* by Rns RS2 region variants

Mutation	%WT[a] *cfaA-lacZ*$^{\Delta 60}$ ± SD	%WT[a] *nlpA*^{cfaA}*-lacZ* ± SD	%WT[b] Protein ± SD	Category[c]
I12G	41±11	65±7	49±13	Less protein
I12E	47±12	70±7	50±0	Less protein
I12W	73±23	47±3	30±11	Less protein
I12R	60±6	63±4	58±3	Less protein
I12S	72±9	77±4	103±32	~WT
K13C	81±10	100±4	153±21	~WT
K13S	83±12	86±4	106±5	~WT
K13G	84±16	78±4	136±25	~WT
I14W	6±2	58±3	104±0	Activate/Bind
N15G	17±3	76±4	64±9	Activate
N15L	23±9	66±2	179±1	Activate
N15D	75±1	88±4	109±15	~WT
N16C	5±2	66±3	72±8	Activate
N16D	25±3	76±3	106±15	Activate
N16G	64±8	83±4	146±20	~WT
I17P	0±0	69±4	93±0	Activate
I17N	2±1	78±3	96±1	Activate
I17R	2±3	87±4	79±20	Activate
I17S	3±4	65±4	92±0	Activate
I17Y	50±	83±3	81±6	Activate
I17V	60±3	77±3	98±1	~WT
I17C	79±4	102±4	91±13	~WT
M18P	14±2	45±5	32±13	Less protein
M18G	68±17	97±4	86±2	~WT
M18R	80±16	67±4	50±8	Less protein

[a]Plasmids (pHG165) encoding wild type and different variants of Rns RS2 region were transformed into *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ* fusion strains and assayed by β-galactosidase reporter gene assay. Expression of *cfaA-lacZ*$^{\Delta 60}$ and repression of *nlpA*^{cfaA}*-lacZ* by wild-type Rns was set to 100%. Results are the average of three independent experiments. Percentage *cfaA-lacZ*$^{\Delta 60}$ expression less than 60% of wild type and non-overlapping standard deviation were used as a significant change in *cfaA-lacZ*$^{\Delta 60}$ expression. Non-overlapping standard deviation and three times the SD of wild type were used as a significant change in *nlpA*^{cfaA}*-lacZ* repression. [b]Rns levels were estimated for different variants by Western blotting and presented as a relative

percentage of wild type. Non-overlapping SD and protein levels less than 60% of wild type were used to determine level of significance

[c]Different categories of mutants isolated: Less protein, variants with decreased Rns protein levels; ~WT, variants with wild type activation, repression and protein levels; Activate, variants with significant defects in transcription activation; Bind, variants with significant defects in DNA binding; Activate/Bind, variants with defects in both transcription activation and DNA binding.

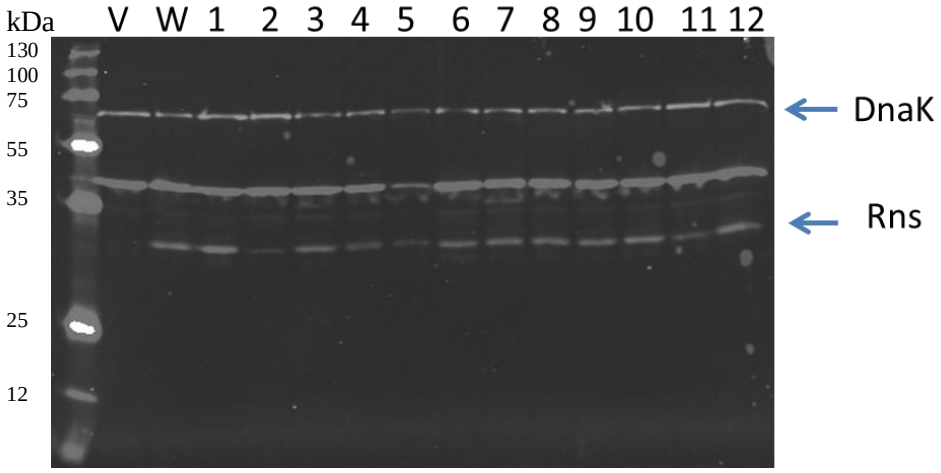


Fig. 9. Western blot analysis of variants in the RS2 region. Western blot was performed to quantify the level of wild-type Rns and different variants of Rns from RS2 region. DnaK served as loading control (Top arrow). The bottom arrow represents the Rns band. Numbers 1 to 12 represent different variants isolated from residues I12, K13, I14 and N15 in the RS2 region. V, vector control; W, wild-type Rns; 1, I12S; 2, I12W; 3, I12R; 4, I12G; 5, I12E; 6, K13S; 7, K13C; 8, K13G; 9, I14W; 10, N15D; 11, N15G; 12, N15L;

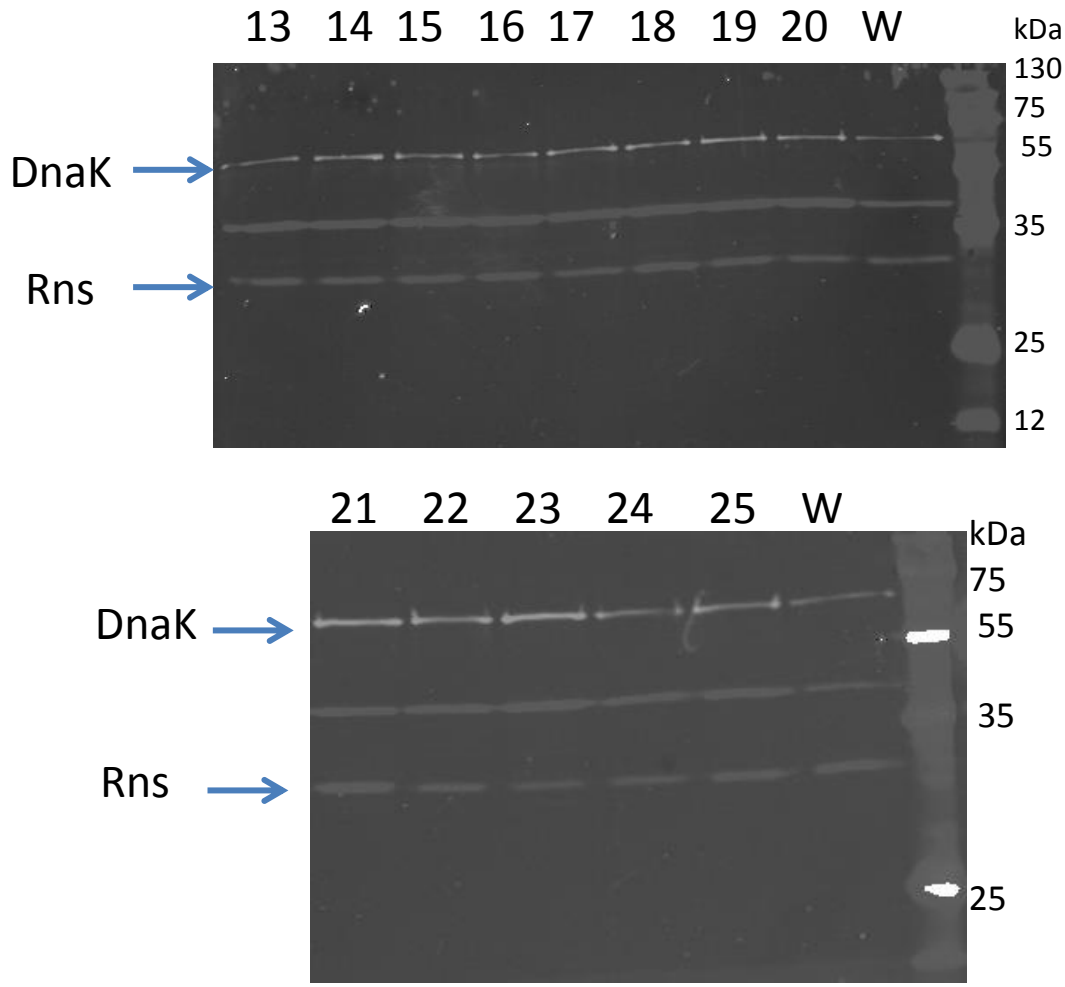


Fig. 10. Western blot analysis of variants in the RS2 region. Western blot is performed to quantify the level of wild-type Rns and different variants of Rns from RS2 region. DnaK served as loading control (Top arrow). The bottom arrow represents the Rns band. Numbers 13 to 20 on the top gel and 21 to 25 on the bottom gel to 25 represents different variants isolated from residues N16, I17 and M18. W, wild-type Rns; 13, N16C; 14, N16D; 15, N16G; 16, I17R; 17, I17P; 18, I17N; 19, I17S; 20, I17Y; 21, I17C; 22, I17V; 23, M18P; 24, M18R; 25, M18G;

Variants in the Rns DBD AS2 region affect Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression. Site-directed random mutagenesis was also performed for 9 residues in the Rns-DBD that are predicted to be contacted by the residues in the RS2 region of NTD (based on I-TASSER prediction). We refer this region as AS2 region (allosteric site in the sub domain 2 of DBD). According to the predicted I-TASSER structural model, the residues in Rns AS2 region were in a similar position to residues in the ToxT structure that were positioned to make inter-domain interactions with the ToxT-NTD (79, 157). The AS2 region constituted residues Q211, S215, K216, L219, K249-G252 and Y261. In the predicted Rns structure, residues Q211, S215, K216 and L219 were in the α7 helix (helix between two helix-turn-helix motifs) and K249-G252 and Y261 were in the α9 – loop - α10 region (Fig. 11). These residues were structurally analogous to residues in the ToxT-DBD (α7 helix and α9 – loop - α10 region) that were positioned to contact the residues in the ToxT-NTD (ToxT Y20 to Y26) (Fig. 12).

Mutagenesis of Rns residues in the α7 helix (Q211, S215, K216, L219) resulted in isolation of 17 variants, among which 13 (77%) had wild type *cfaA-lacZ*$^{\Delta 60}$ expression (Table 6) and four were defective for *cfaA-lacZ*$^{\Delta 60}$ expression. Among the four variants that were defective, two of them (Q211P, S215V) were defective for protein levels and thereby not considered as true defects of *cfaA-lacZ*$^{\Delta 60}$ expression. The other two defective variants at residue K216 (K216P, K216L) were the only two variants isolated at this position. Both of these variants had near wild-type protein levels, suggesting that Rns K216 is important for Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression.

We also mutagenized 5 positions in the predicted α9 – loop - α10 region of Rns AS2, and isolated a total of 24 unique variants. Unlike the residues in the α7 region of the AS2 region, the majority of the AS2 variants in the α9 – loop - α10 region were defective for Rns-dependent

cfaA-lacZ$^{\Delta 60}$ expression (75% of variants, 18 out of 24) (Table 6). Our Western blots showed no detectable changes in protein levels for these variants (Table 6). Strikingly, all 14 variants isolated at residues H250, Y251 and G252 had reduced Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression, suggesting the importance of these consecutive residues in Rns function (Table 6). Two of the five variants isolated at K249 and two of the six variants isolated at Y261 were defective for Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression (Table 6). Given that the majority of variants at K249 and Y261 had *cfaA-lacZ*$^{\Delta 60}$ expression similar to wild type, the defective variants may be gain of function mutants. This suggests that residues K249 and Y261 may not be required for *cfaA-lacZ*$^{\Delta 60}$ expression or that specific substitutions can influence other residues that play a direct role in *cfaA-lacZ*$^{\Delta 60}$ expression. Altogether, our results suggest that residues H250, Y251 and G252 in the $\alpha 9$ – loop - $\alpha 10$ region are important for Rns-dependent *cfaA-lacZ*$^{\Delta 60}$ expression.

Distinguishing DNA binding defects from transcription activation defects in AS2 region.

Similar to the previous analysis of RS2 region variants in the repression strain (*nlpA*^{cfaA}-*lacZ*), we tested variants isolated in the AS2 region. Comparison of variants for *nlpA*^{cfaA}-*lacZ* repression, *cfaA-lacZ*$^{\Delta 60}$ expression, and protein levels were used to identify the variants that impacted DNA binding or transcription activation. Our western blot analysis showed that none of the variants had defects in the protein levels. Among the two variants isolated at position K216, one variant K216P was defective exclusively for *cfaA-lacZ*$^{\Delta 60}$ expression, suggesting that this variant is defective for transcription activation. The other variant K216L was defective for both *cfaA-lacZ*$^{\Delta 60}$ expression and *nlpA*^{cfaA}-*lacZ* repression (defects in both DNA binding and transcription activation). However, K216L had ~50 fold higher defects in *cfaA-lacZ*$^{\Delta 60}$ expression than *nlpA*^{cfaA}-*lacZ* repression, suggesting that the variant is mainly defective for transcription activation function of Rns.

A total of five variants were isolated from the residue H250. Among the five variants, two were defective only for *cfaA-lacZ*$^{\Delta 60}$ expression (H250R, H250V), suggesting that these variants show defects in transcription activation. Three variants were defective for both *cfaA-lacZ*$^{\Delta 60}$ expression and *nlpA*^{cfaA}*-lacZ* repression (H250I, H250N and H250T), suggesting that these variants show defects in DNA binding and transcription activation. Among these three variants, one variant (H250I) had ~6 fold higher defects in *cfaA-lacZ*$^{\Delta 60}$ expression than *nlpA*^{cfaA}*-lacZ* repression. Overall, the residue H250 seems to be required for both DNA binding and transcription activation (Table 6). All the four variants isolated from AS2 region residue Y251 were defective for *cfaA-lacZ*$^{\Delta 60}$ expression activation but not for repression, indicating that this residue is important for transcription activation function of Rns (Table 6). For residue G252, three variants (G252Q, G252S and G252A) were defective for *cfaA-lacZ*$^{\Delta 60}$ expression and two variants (G252N and G252R) were defective for both *cfaA-lacZ*$^{\Delta 60}$ expression and *nlpA*^{cfaA}*-lacZ* repression. G252R showed ~7 fold higher defects in *cfaA-lacZ*$^{\Delta 60}$ expression than *nlpA*^{cfaA}*-lacZ* repression. Overall, this residue seems to be required for both DNA binding and transcription activation.

For residue K249, one variant (K249Y) showed defects in *cfaA-lacZ*$^{\Delta 60}$ expression and another one (K249A) showed defects in both *cfaA-lacZ*$^{\Delta 60}$ expression and *nlpA*^{cfaA}*-lacZ* repression. For residue Y261, two variants showed only *cfaA-lacZ*$^{\Delta 60}$ expression defects (Y261G and Y261E) (Table 6). However, majority of the variants at K249 and Y261 were found to have no *cfaA-lacZ*$^{\Delta 60}$ expression and *nlpA*^{cfaA}*-lacZ* repression defects, likely suggesting that the defective variants at these positions may be gain of function variants (Table 6).

Overall, our analysis of variants in both *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ* strains suggested that two residues in AS2 region (K216 and Y251) were required for transcription activation

function of Rns and two residues H250 and G252 were required for both DNA binding and transcription activation function of Rns.

Analysis of Rns I14T and N16D variants in *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ* fusion strains.Previous reports by Basturea *et al* showed that Rns I14T and N16D variants (expressed from a high copy number plasmid pUC18) alter the Rns DNA sequence specificity at *rns* promoter region, and thereby decrease the expression from *rns-lacZ* fusion (156). Their results also showed that I14T did not affect *cfaA-lacZ* and *nlpA-lacZ* expression (156) and thus suggesting that I14T substitution may not impact the DNA binding by Rns at *cfaA* and *nlpA* promoter regions, and transcription activation from *cfaA* promoter region. Results of N16D substitution showed greater defects in *cfaA-lacZ* expression than *nlpA-lacZ* expression (156), and thus indicated us that N16D may impact the transcription activation function by Rns from *cfaA* promoter region. Our mutagenesis of I14 residue did not result in isolation of I14T substitution. However, we isolated N16D substitution that was defective for transcription activation function by Rns from *cfaA* promoter region. Given that our *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ* fusions differ from *cfaA-lacZ* (differs in number of Rns binding sites) and *nlpA-lacZ* (differs in both number and sequence of binding sites) fusions used by Basturea et al, we tested the effect of Rns I14T and N16D (expressed from high copy number plasmid pUC18) on *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ* expression. Our analysis showed that I14T was not defective for either *cfaA-lacZ*$^{\Delta 60}$ or *nlpA*^{cfaA}*-lacZ* expression, and thus indicating that I14T may not interfere with DNA binding and transcription activation function by Rns, at least at the Rns binding site from *cfaA* (Fig. 13). Our analysis of N16D (expressed from pUC18 plasmid) for *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ* expression showed that N16D was defective only for *cfaA-lacZ*$^{\Delta 60}$ expression (Fig. 13), and thus indicating defects in transcription activation function from *cfaA* promoter region. The impairment of

transcription activation function by N16D further supports our observation that the residue N16 may be required for transcription activation function by Rns.

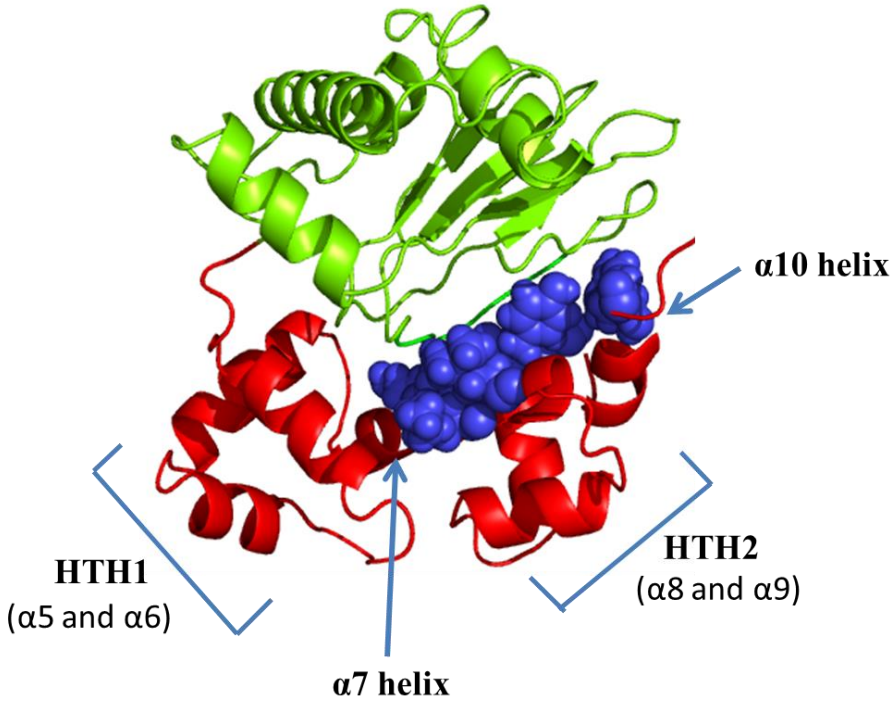


Fig. 11. Predicted Rns three dimensional structure with AS2 region mapped. I-TASSER (157) was used to predict the 3D structure of the Rns. NTD was shown in green and DBD was shown in red. Residues corresponding to the RS2 region in the DBD were shown as blue spheres.

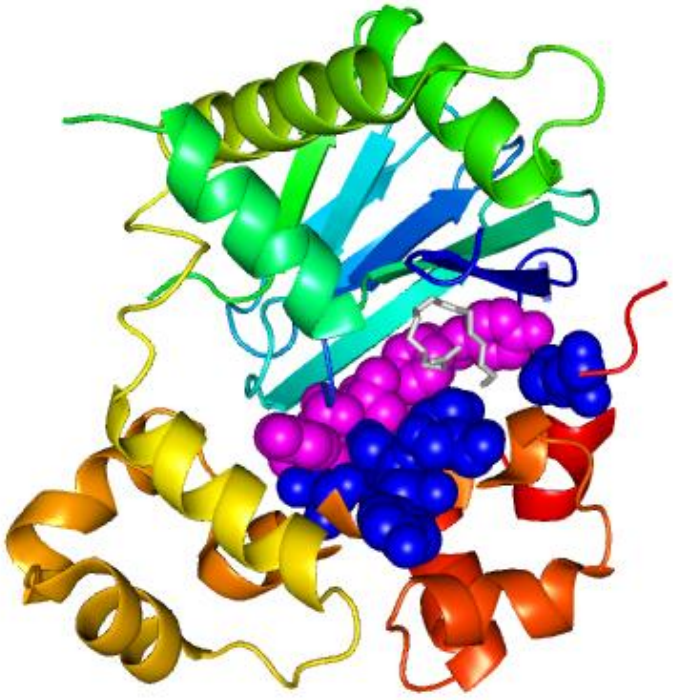


Fig. 12. ToxT structure with residues corresponding to RS2 and AS2 regions mapped (79)**.** Ribbon diagram of the ToxT with the effector palmitoleic acid (grey, shown in stick form) bound in the NTD. The ToxT structure is colored with rainbow effect with the violet at the N-terminus and red at the C-terminus. Residues shown as purple spheres indicate the ToxT residues that correspond to Rns RS2 region residues. Residues shown as blue spheres indicate the ToxT residues that correspond to Rns AS2 region residues.

Table 6. Transcription activation of *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ* by Rns AS2 region variants.

Mutation	%WT[a] *cfaA-lacZ*$^{\Delta 60}$±SD	%WT[a] *nlpA*^{cfaA} *lacZ*±SD	%WT[b] Protein±SD	Category[c]
Q211P	15±6	40±4	35±12	Less protein
Q211N	67±16	83±3	140±1	~WT
Q211L	69±10	95±4	104±0	~WT
Q211S	77±15	87±3	ND[d]	~WT
Q211G	96±11	97±4	106±16	~WT
S215V	46±17	55±10	39±17	Less protein
S215D	74±8	98±4	93±10	~WT
S215G	84±3	106±5	ND[d]	~WT
S215K	113±7	102±4	93±11	~WT
K216L	1±0	57±4	83±25	Activate/Bind
K216P	9±1	86±4	78±4	Activate
L219H	95±9	78±4	89±0	~WT
L219I	97±8	65±2	105±6	~WT
L219A	107±18	72±5	79±12	~WT
L219K	122±11	74±1	113±19	~WT
L219M	121±9	78±4	92±1	~WT
L219Y	124±22	80±3	96±0	~WT
K249Y	31±8	81±4	93±11	Activate
K249A	45±8	52±8	79±10	Bind
K249N	65±12	79±4	95±7	~WT
K249E	79±3	75±4	131±27	~WT
K249R	107±21	91±4	91±13	~WT
H250I	9±3	49±6	128±6	Activate/Bind
H250V	18±11	81±9	122±28	Activate
H250R	42±8	81±5	104±1	Activate
H250N	44±13	64±6	100±29	Bind
H250T	48±12	62±6	92±28	Bind
Y251L	1±0	68±6	140±43	Activate
Y251V	1±0	78±5	68±7	Activate
Y251T	2±1	71±4	94±9	Activate
Y251F	19±6	71±5	71±2	Activate
G252R	9±8	62±3	113±25	Activate/Bind
G252Q	11±10	79±4	94±33	Activate
G252A	17±12	70±4	119±47	Activate
G252S	19±14	82±5	78±3	Activate
G252N	43±12	58±4	86±12	Bind
Y261E	9±7	78±4	92±11	Activate
Y261G	20±14	68±4	111±28	Activate
Y261C	64±10	76±3	85±31	~WT
Y261M	64±6	83±3	140±13	~WT

Y261A	68±9	64±3	118±23	~WT
Y261R	68±11	76±3	74±2	~WT

[a]Plasmids (pHG165) encoding wild type and different variants of Rns AS2 region were transformed into *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ* fusion strains and assayed by β-galactosidase reporter gene assay. Expression of *cfaA-lacZ*$^{\Delta 60}$ and repression of *nlpA*^{cfaA}*-lacZ* by wild-type Rns was set to 100%. Results are the average of three independent experiments. Percentage *cfaA-lacZ*$^{\Delta 60}$ expression less than 60% of wild type and non-overlapping standard deviation were used as a significant change in *cfaA-lacZ*$^{\Delta 60}$ expression. Non-overlapping standard deviation and three times the SD of wild type were used as a significant change in *nlpA*^{cfaA}*-lacZ* repression. [b]Rns levels were estimated for different variants by Western blotting and presented as a relative percentage of wild type. Non-overlapping SD and protein levels less than 60% of wild type were used to determine level of significance

[c]Different categories of mutants isolated: Less protein, variants with decreased Rns protein levels; ~WT, variants with wild type activation, repression and protein levels; Activate, variants with significant defects in transcription activation; Bind, variants with significant defects in DNA binding; Activate/Bind, variants with defects in both transcription activation and DNA binding.

[d] ND – protein levels not determined

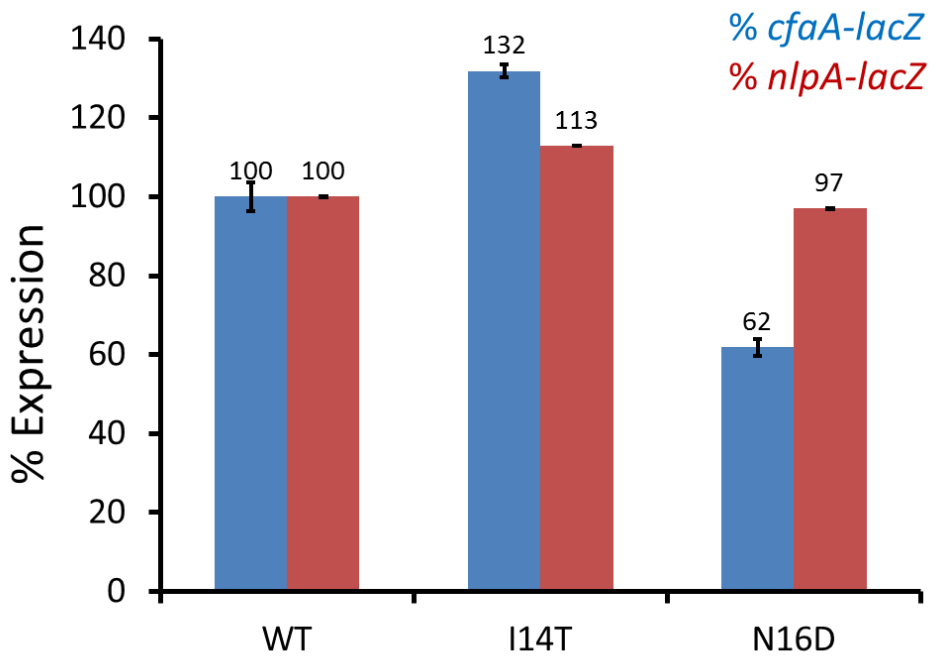


Fig.13. Analysis of Rns variants I14T and N16D (expressed from pUC18) from *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ*. Rns variants I14T and N16D expressed from a high copy number plasmid pUC18 were analyzed for expression from *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ* fusion strains using β-galactosidase assays. Results from both these strains were compared to identify the defects in DNA binding or transcription activation function. Expression from *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ* by wild type Rns were adjusted to 100%. Results are the average of three independent experiments and error bars represent the standard deviations of the mean.

Chapter 4

Purification and properties of Rns

In order to perform in vitro assays and to generate antibodies for Western blotting we wanted to have purified Rns protein. In addition, we wanted to test oligomerization state of Rns (monomer or dimer), as previous reports on oligomerization state of Rns were quite contradictory (155, 156). Progress in biochemical analysis of AraC family proteins has traditionally been hampered by poor solubility and stability of these proteins *in vitro*. Therefore, in order to increase the solubility of Rns, I made fusion constructs that links Rns with proven solubility enhancement proteins GB1 (~8 kDa) and MBP (~42.5 kDa). Studies by Basturea *et al* had shown that Rns can be successfully purified using MBP tag (156). Rns has many rare codons whose expression is limited in *E. coli* strains (most of the rare codons were present in the NTD). In order to overcome the limitations of the rare codons, codon-optimized *rns* (purchased from GenScript, Piscataway, NJ) was linked to the GB1 tag. Table 7 shows the list of Rns constructs made for overexpression.

Despite my efforts in overexpression of Rns and Rns-NTD with GB1 attached at both the N-terminus and C-terminus of the protein (expressed from *lacI* controlled T7 promoter), I was not able to detect any overexpression from these constructs. However, one construct Rns-DBD-GB1-His6 overexpressed in all the conditions. In addition, full-length Rns and Rns-NTD overexpressed when linked with MBP tag and expressed from P*tac* promoter (*lacIq* controlled).

Purification of Rns-DBD-GB1-His6. Rns-DBD-GB1-His6 was overexpressed at 37^{o}C from pDZ3 plasmid from the strain BL21 (DE3)DnaY (100 ml initial culture). Cells were lysed by sonication and run on SDS-PAGE which showed that the protein was insoluble. The pellet fraction was resuspended in buffer containing either 2 or 4 or 6 M urea and incubated overnight. Protein was found to be soluble by ~50% in 6M urea. The soluble fraction of the protein was purified using Ni^{2+} affinity purification under denaturing conditions (6 M urea containing elution buffer). Urea was subsequently removed from the sample using a HiPrep 26/10 desalting column

into 10ml of elution buffer (20 mM Tris-base, 500 mM NaCl, 500 mM Imidazole, 0.1 mM TCEP, pH 7.9) (Fig. 14A). The protein was stable at 0.3 mg/ml.

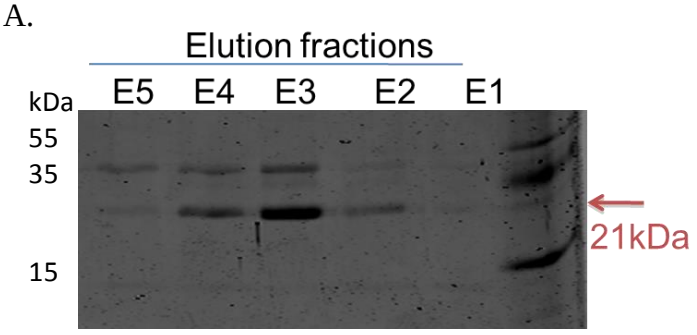
Alternatively, Rns-DBD-GB1-His6 was soluble (~50%) when overexpressed overnight at 15^{o}C. The soluble fraction of the protein was purified using Ni^{2+} affinity purification into 10 ml elution buffer (Fig. 14B). The protein was stable at 1.1 mg/ml. Protein fractions showing only the Rns-DBD-GB1-His6 were pooled and tested for binding using electrophoretic mobility shift assays (EMSAs). The result in the figure 15 demonstrates that Rns-DBD-GB1-His6 could bind to the DNA and the amount of binding decreases with decrease in protein concentrations. For long-term storage of active Rns-DBD-GB1-His6, glycerol was added to a final concentration of 25% and stored at -80^{o}C.

Previously, DNaseI footprinting experiments by Basturea *et al* (156) using a longer MBP-Rns construct (Rns residues 128-265) that included the DBD has shown that this construct did not bind to the *cfaA* promoter region even at high protein concentrations of 2.5 µM. Surprisingly, our EMSA experiments with the purified Rns-DBD-GB1-His6 showed that this protein can bind to *cfaA* promoter region even at concentrations of 1 µM. However, even at highest protein concentration tested (9 µM), only 7% of the DNA was bound by the protein. We expect that the differences in the sensitivities of the techniques may have contributed to observation of these varied results. DNase foot printing is more sensitive than EMSA in terms of using P^{32} labeled DNA as opposed to IR labeled DNA in our EMSAs. However, it is harder to detect 7% DNA binding in DNase foot printing as we look for decrease in intensity of the DNA band by 7% in footprinting (100% to 93%) compared to gaining a new band in EMSA (0% to 7%).

Table 7. Rns constructs for overexpression and purification.

Construct	Cloning	Overexpression
GB1-Rns	Yes	None
Rns-GB1	Yes	None
GB1-RnsNTD	Yes	None
RnsNTD-GB1	Yes	None
RnsDBD-GB1	Yes	Yes
GB1-Rns^{co}	Yes	None
Rns^{co}-GB1	Yes	None
MBP-Rns	Yes	Yes
MBP-RnsNTD	Yes	Yes
MBP-RnsDBD	Yes	Yes

A.



B.

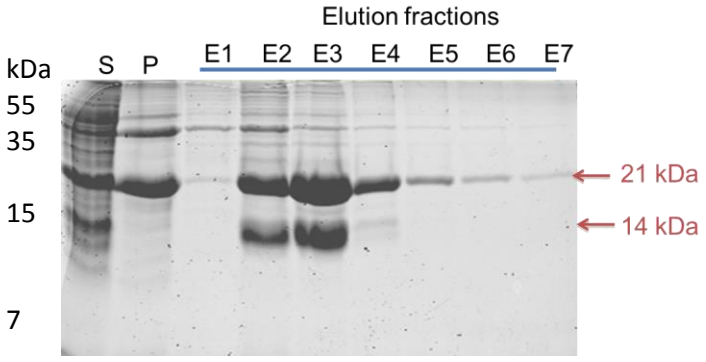


Fig. 14. Purification of Rns-DBD-GB1-His6. A) Fusion protein was solubilized from the pellet using urea and purified using Ni^{2+} affinity chromatography into elution buffer (20 mM Tris-base, 500 mM NaCl, 500 mM Imidazole, 0.1 mM TCEP, pH 7.9) containing 6M urea. **B)** Purification of fusion protein from soluble fraction by using Ni^{2+} affinity chromatography into elution buffer. S, supernatant fraction; P, pellet fraction; E1-E7, elution fractions. In addition to Rns-DBD-GB1-His6 (22 kDa), an additional prominent lower band of ~14 kDa was found in few fractions consistently, although further experiments were not carried out to identify the protein representing the lower band.

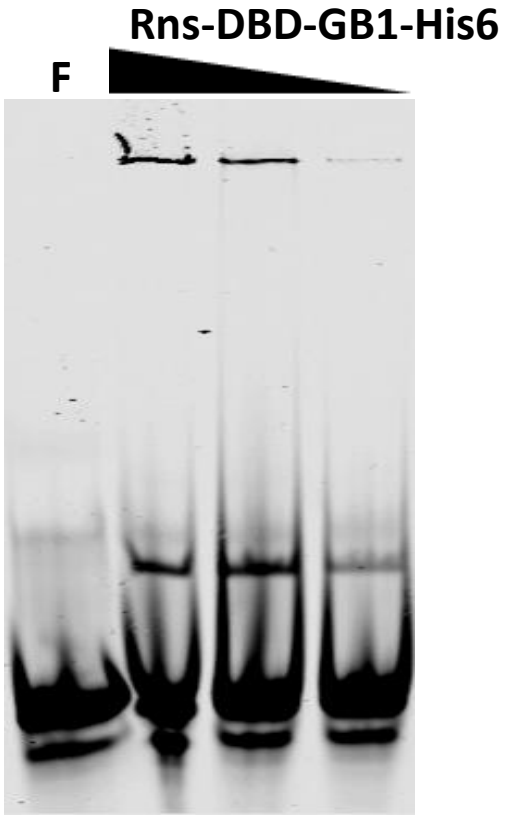


Fig. 15. Electrophoretic mobility shift assay of DNA binding by Rns-DBD-GB1-His6. Binding of purified Rns-DBD-GB1-His6 to a DNA fragment containing the Rns binding site from the *cfaA* promoter was assayed using electrophoretic mobility shift assays (EMSAs). Black triangle represents decreasing concentrations of Rns-DBD-GB1-His6, 9 μM to 1μM, with serial three-fold dilutions. DNA was added at a final concentration of 0.11 μM.

Overexpression and Purification of MBP-Rns protein. MBP-Rns fusion protein was expressed from pMBP-Rns1 plasmid in the *E. coli* strain KS1000. The strain was also transformed with pRARE2 (Novagen) plasmid that has seven tRNAs genes that encode rare codons required by Rns. After overnight induction, cells were sonicated and centrifuged to remove the cell debris. The soluble supernatant was then loaded on a BioLogic LP chromatography system to an amylose resin column and eluted in 25 ml of elution buffer containing 10 mM maltose [20 mM Tris-base, 500 mM NaCl, 1 mM EDTA, 0.1 mM TCEP, 15% Glycerol, 10 mM Maltose pH 7.6] (Fig. 16). Although the final protein had some impurities, we obtained a total of ~5 mg of soluble protein from 1 L initial culture. To further purify the protein and to determine whether Rns act as monomer or dimer, eluted fractions were concentrated to not more than 2 mg/ml (above which it aggregated) and passed through a size exclusion Sephacryl 16/60 S-200 chromatography column. The majority of the protein eluted at a mass of very large aggregates, although some fraction of the protein was found at the size of monomer (~73 kDa) (Fig. 17). In addition, some fraction of the protein was eluted at a size of MBP (~40 kDa) indicating the partial cleavage of the MBP from the Rns protein. Protein markers of sizes 150 kDa and 75 kDa corresponding to dimer and monomer of Rns, respectively were run through column and used to determine the oligomerization status of Rns fractions.

In order to find the suitable buffer in which Rns is less prone to aggregation, I exchanged the subsequent protein preps in to a simple Tris-buffer with less salt and no maltose (15 mM Tris-base, 75 mM KCl, 5 mM β-mercaptoethanol, 10% Glycerol, pH 7.6) using a Hiprep 26/10 desalting column. The protein was concentrated to ~2 mg/ml and then passed through a size exclusion chromatography column. Unlike elution buffer, protein did not visibly aggregate in this buffer. Majority of the protein was eluted at ~53 ml which corresponds to a molecular

weight of ~75 kDa in this buffer (Fig. 18) (Running of a protein marker of 75 kDa showed that 53 ml corresponds to ~75 kDa in this buffer). This suggests the MBP-Rns behaves as a monomer in solution *in vitro*. Similar to the previous observation, some fraction of the protein was eluted at a size of MBP (~40 kDa) indicating the partial cleavage of the MBP from the Rns protein. EMSAs were performed using purified MBP-Rns protein to test whether the protein is active *in vitro.* DNA fragments (one IR labeled and one non-IR labeled oligo) containing the Rns binding site from *cfaA* promoter region was used to test *in vitro* DNA binding ability of Rns. EMSA results indicated that MBP-Rns is active *in vitro* and can bind to DNA (Fig. 19). In addition, presence of large MBP tag did not inhibit the binding of protein to the DNA, suggesting that MBP tag may not interfere with Rns. This observation further supports the previous observations that MBP tag does not interfere with Rns activity *in vivo* and *in vitro* (147, 154). Purified active MBP-Rns was stored at -80^{o}C for long-term storage.

The purified MBP-Rns was injected into rabbit five times (at 15 days interval) to make polyclonal Rns antibodies for the use in Western blotting (Cocalico Biologicals, Reamstown, PA). The sera obtained were affinity purified against purified MBP-Rns protein to isolate Rns antibodies and tested for their ability to detect Rns expressed from pHG165 in the *E. coli* strain SME4336. Figure 20 demonstrates that the affinity purified Rns antibodies detect Rns from SME4336, with very little background.

Rns was found to be very prone to aggregation when cleaved from MBP tag. Cleavage reaction were performed at room temperatures (for 1-4 h) using a protease factor XA (1:100 ratio of XA to MBP-Rns) that cleaves in the linker region between MBP and Rns. When the cleavage reaction components were run on SDS-PAGE, only a high intense band that corresponds to MBP

protein was visible (Fig. 21). It might be possible that majority of Rns aggregated in the tube after cleavage. Alternatively, protease factor XA may have cleaved Rns at multiple places.

Given that Rns aggregated when cleaved from MBP protein, crystallization trial were conducted using MBP-Rns fusion protein. MBP-Rns protein was concentrated using a Millipore centricon (50 kDa cut off) and concentrated to ~20 mg/ml (in buffer with 15 mM Tris-base, 75 mM KCl, 5 mM β-mercaptoethanol, 10% Glycerol, pH 7.6) without any visible aggregation. Aggregation at 20 mg/ml concentration was tested using dynamic light scattering (DLS). DLS measures the fluctuations in scattered light intensity due to diffusing particles, and thereby calculates the mean effective diameter of the particles in the population. If the protein is monodisperse, a single mean effective diameter of the protein was possible. Depending on the size of the protein the mean effective diameter of the protein may vary However, if the protein is aggregated we can see measurements of multiple diameters. Our results showed that the majority of the protein scattered belongs to a single diameter and thus indicates no aggregation (Fig. 22). Crystallization trials were conducted using 20 mg/ml concentrated protein. Almost 400 different commercially available crystallization conditions were tested, but none resulted in crystals. In the fusion protein, MBP and Rns were connected by a flexible 25 amino acid linker region (S_3N_{10}LGIEGRISEFGS) that may have resulted in heterogeneity leading to non-formation of crystals. In addition, the predicted secondary structure of Rns showed an unstructured region near the N-terminus of the protein which may have further increased the flexibility between proteins. Future crystallization trials by removing these flexible regions will be needed to test this hypothesis.

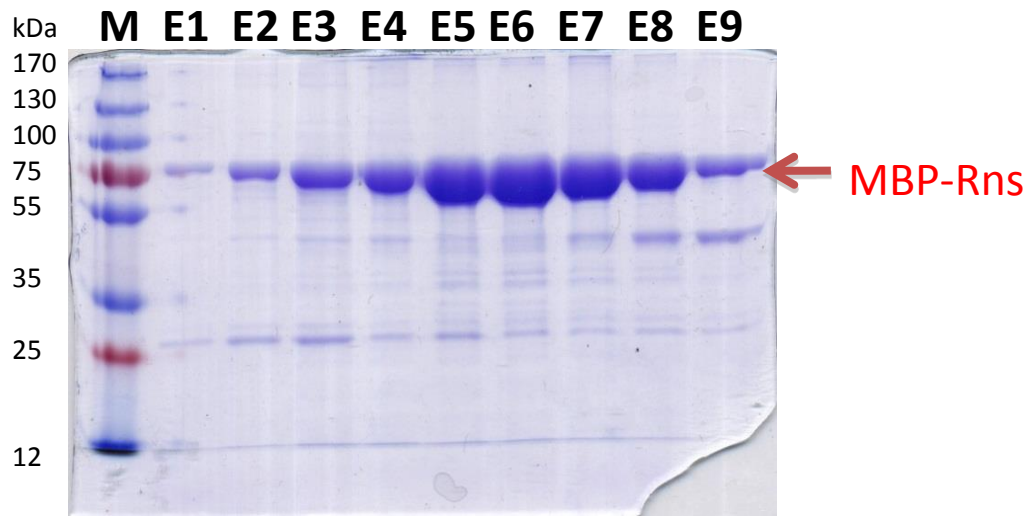


Fig. 16. Purification of MBP-Rns. The fusion protein was purified using amylose affinity chromatography into elution buffer [20 mM Tris-base, 500 mM NaCl, 1 mM EDTA, 0.1 mM TCEP, 15% Glycerol, 10 mM Maltose pH 7.6]. Purified elution fractions were electrophoresed on a 12% SDS-polyacrylamide gel and stained with Coomassie Blue. E1-E9, elution fractions.

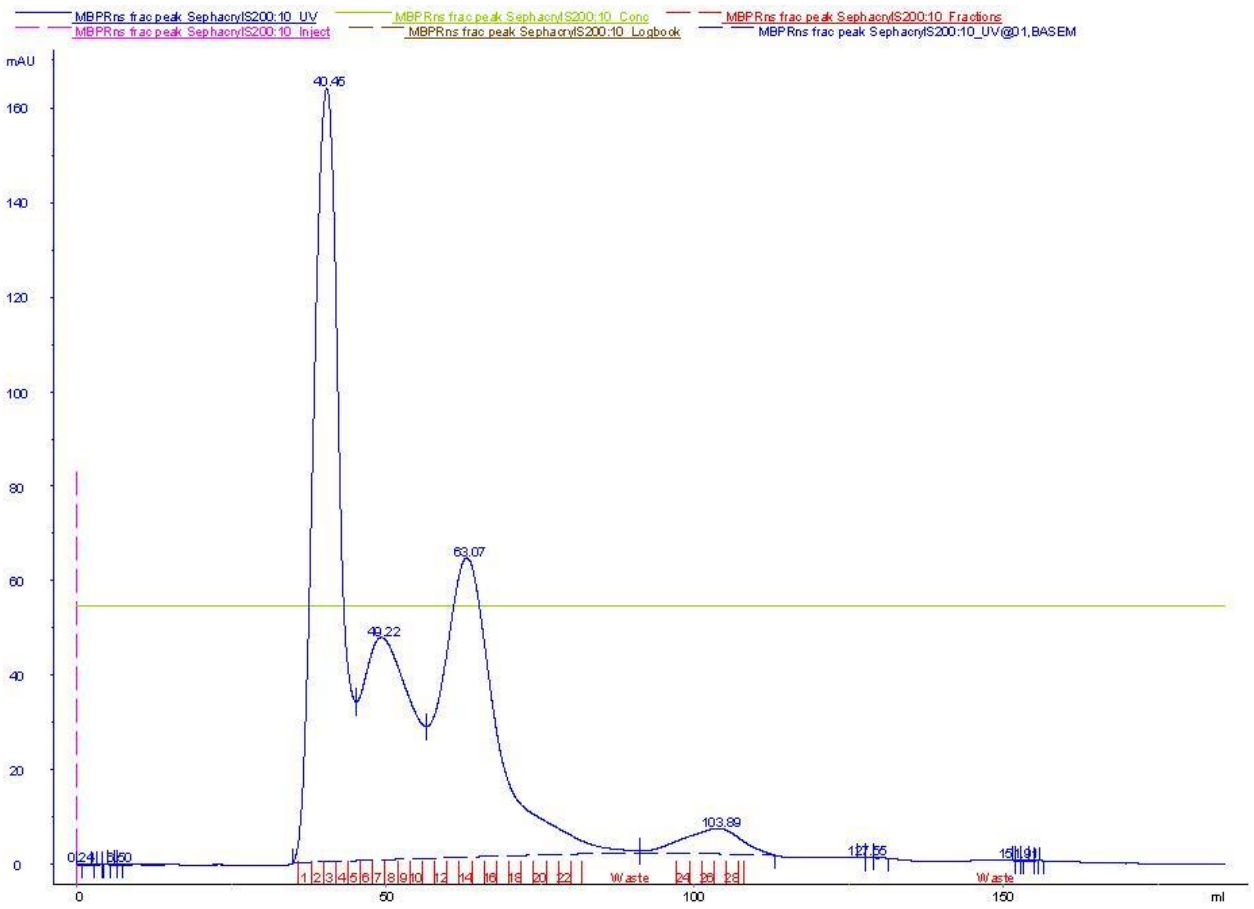


Fig. 17. Size exclusion chromatography of MBP-Rns in elution buffer. Sephacryl 16/60 S-200 chromatography column was used. Y-axis shows UV absorbance at 280 nm and X-axis shows elution volume. The peak at 40.46 ml retention represents aggregation. The next small peak at 49.22 ml retention eluted at the size of ~85 kDa (Probably Monomer). The next peak at 63.07 ml retention was eluted at the size of ~50 kDa.

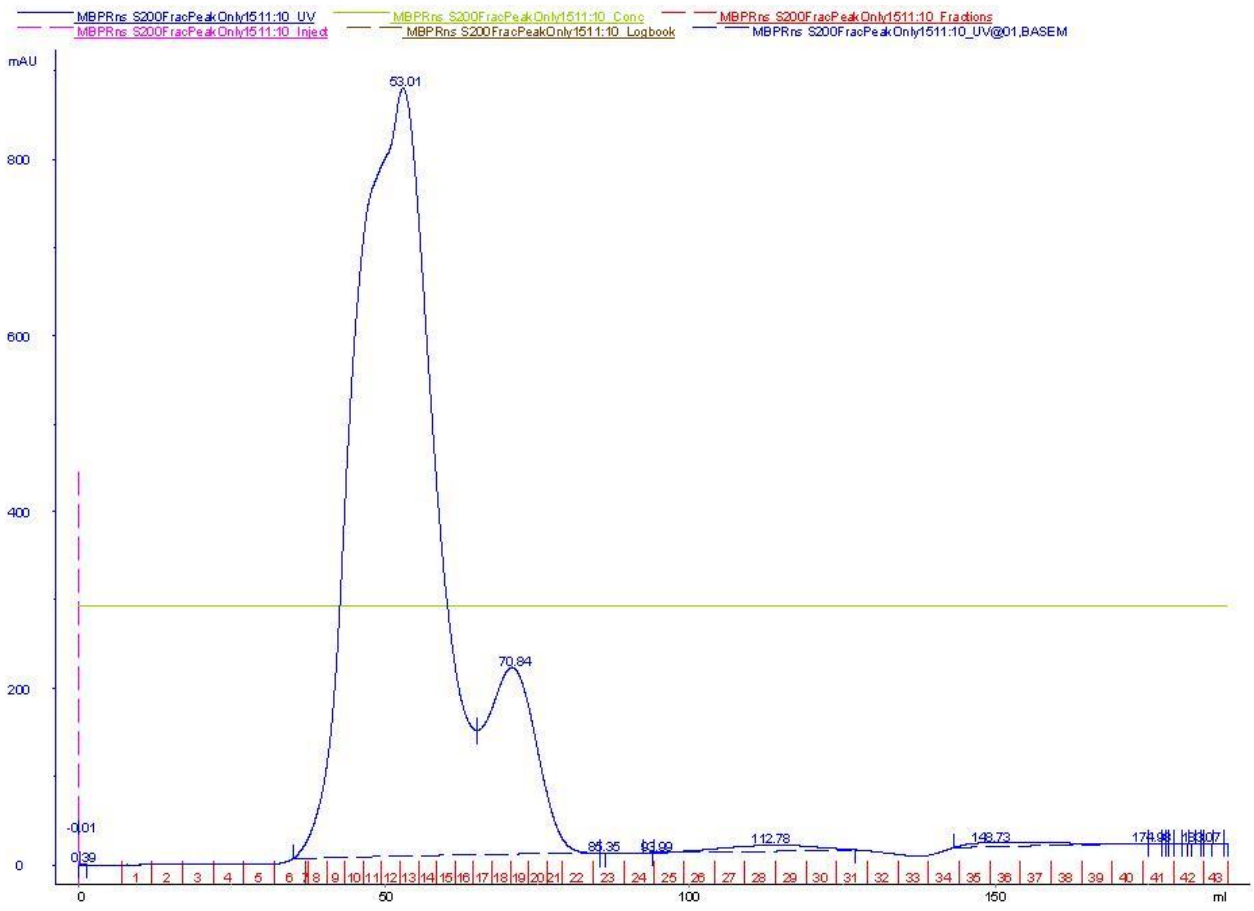


Fig. 18. Size exclusion chromatography of MBP-Rns in elution buffer. Sephacryl 16/60 S-200 chromatography column was used. Y-axis shows UV absorbance at 280 nm and X-axis shows elution volume. The highest peak at 53.01 ml retention eluted at the size of monomer (~75 kDa) (Probably Monomer). The next peak at 70.94 ml retention was eluted at the size of MBP (~40 kDa).

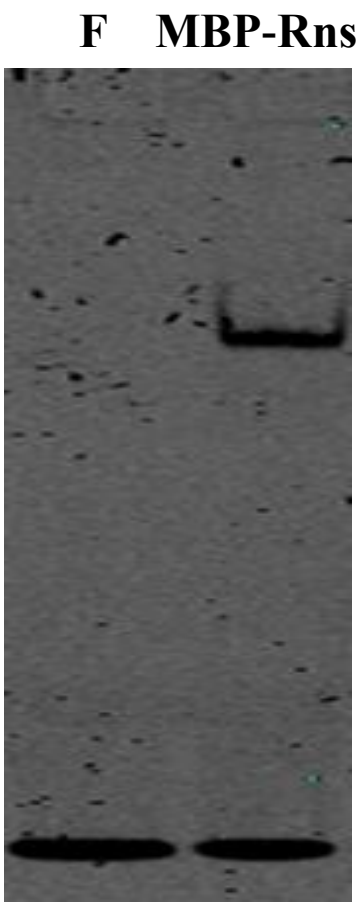


Fig. 19. Electrophoretic mobility shift assay of DNA binding by MBP-Rns. Binding of purified MBP-Rns to a DNA fragment containing the Rns binding site from the *cfaA* promoter was assayed using electrophoretic mobility shift assays (EMSAs). DNA and protein were added at final concentrations of ~2 nM and 0.3 μM, respectively.

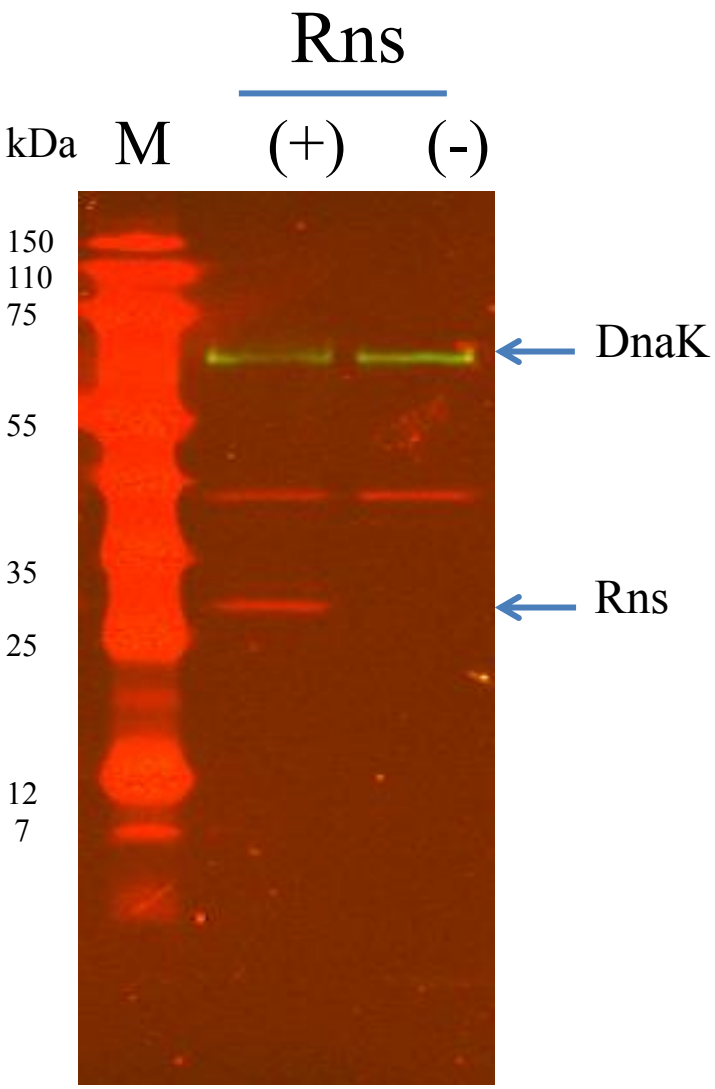


Fig.20. Western blot analysis of affinity purified MBP-Rns antibodies. Affinity purified antibodies were tested in Western blot for their ability to properly recognize Rns expressed from pHG165 in the strain SME4336 (labeled "+"). The strain SME4337 (pHG165 vector control) was used as negative control (labeled "-"). DnaK served as loading control (Top arrow). The bottom arrow represents the Rns band. M, marker.

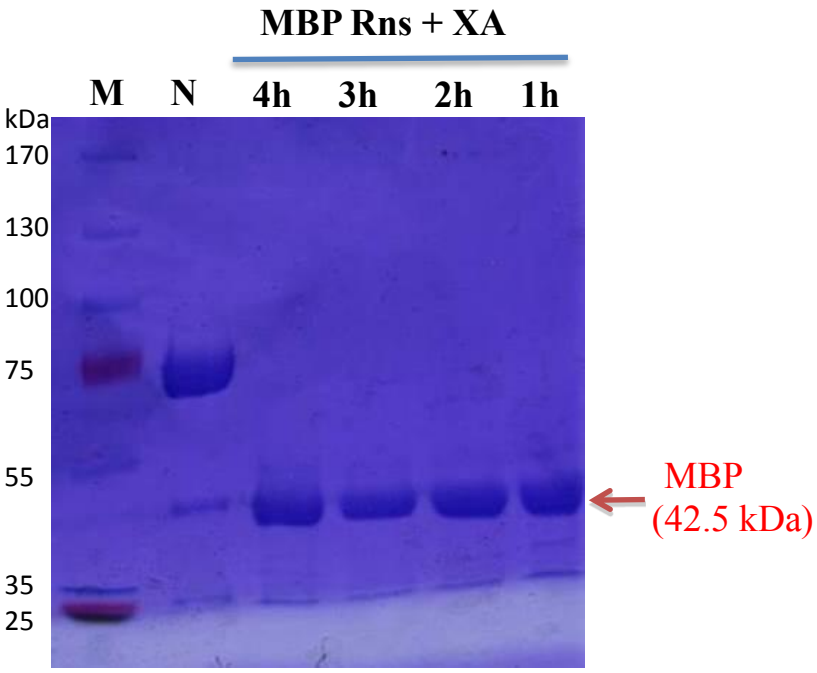


Fig.21. Cleavage of MBP-Rns with protease factor XA. MBP-Rns was incubated with protease factor XA (1:100 dilution) at room temperature for 1-4 h, ran on SDS-PAGE gel, and stained with Coomassie Blue. M, Marker; N, negative control (MBP-Rns without factor XA added).

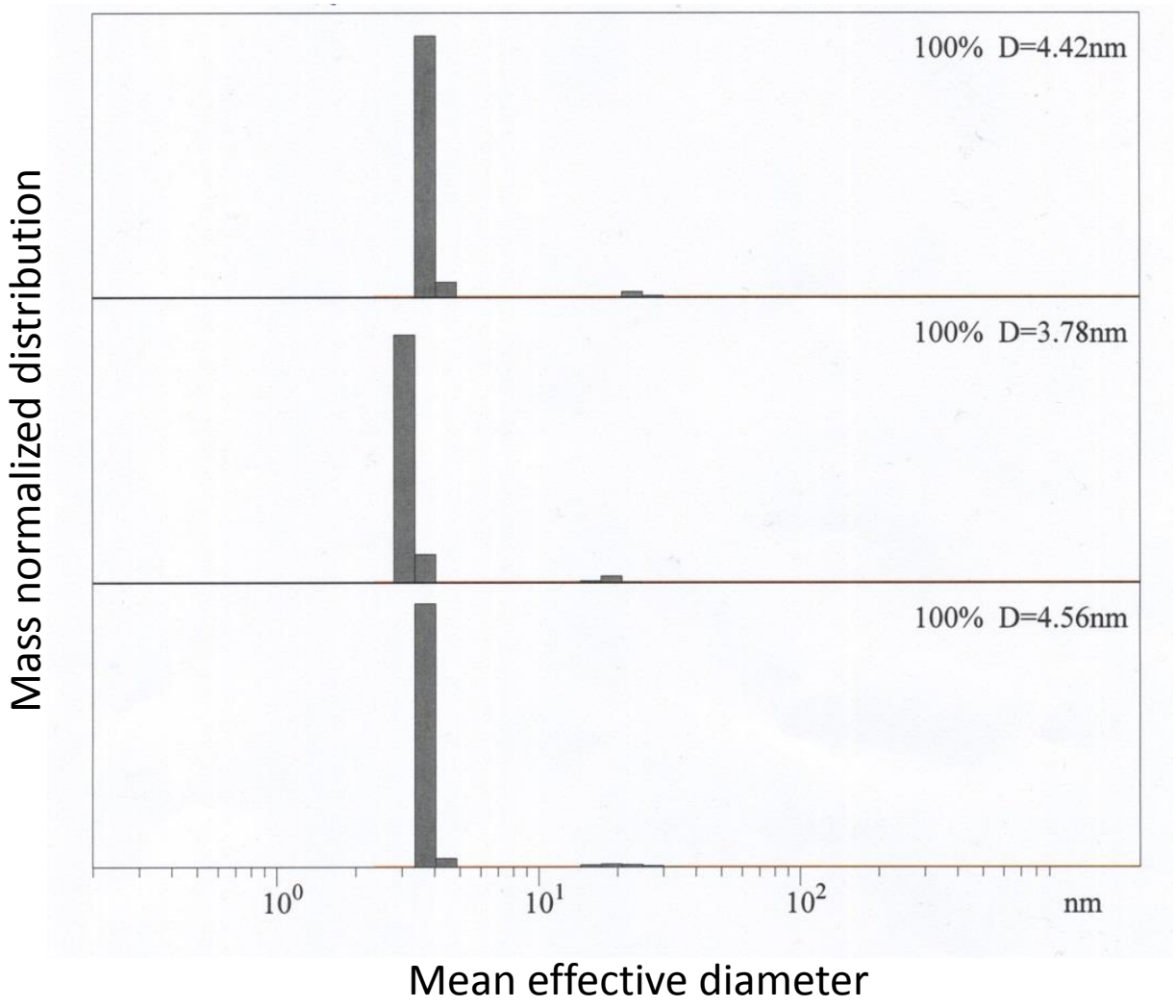


Fig.22. Dynamic light scattering analysis of MBP-Rns protein. A 200 µl of 20 mg/ml concentrated protein was analyzed to test the aggregation levels of the protein. Each of the three panels represents an average of 60 different readings taken in a span of 60 sec. D, mean radius of the protein molecules in solution.

Chapter 5

Small molecule inhibitor of AraC family proteins Rns and VirF

Rns and VirF are the virulence regulators in Enterotoxigenic *E. coli* (ETEC) and *Shigella* respectively and are required by their respective pathogens to cause disease (41, 42, 45-49). Rns and its orthologs are required for expression of majority of the colonization factor antigens (CFs) that are required for the successful adherence of ETEC to its host intestinal epithelial cells (41, 42). ETEC causes diarrhea and is responsible for a high rate of mortality in infants and children under the age of five (300,000 – 500,000 deaths annually), especially in developing countries (117, 214). In the United States, ETEC is most often acquired during travel to nations where it is endemic or through sporadic food borne outbreaks (119, 120). VirF is required for the expression of genes responsible for intracellular invasion and cell-to-cell spread, and is thereby central to *Shigella* virulence (45-49). *Shigella* is responsible for 165 million illnesses and over 1.1 million deaths worldwide each year (158, 159).

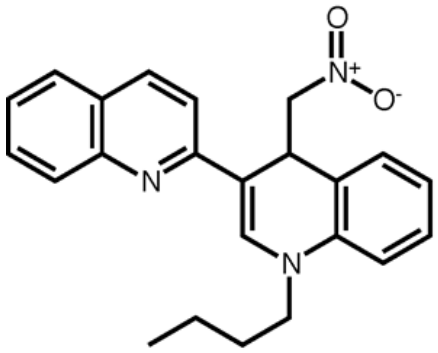
Given the enormous number of infections caused by both these bacteria, it is important to find effective treatment strategies to successfully treat these infections. Despite many efforts there was no vaccine released for public use (although some vaccines are under trials) (215-218). A major obstacle in vaccine development is the substantial heterogeneity of surface antigens among different strains of these pathogens (215-218). Both *Shigella* and ETEC are becoming increasingly resistant to the currently available antibiotics (3, 219). A study carried out by National Antimicrobial Resistance Monitoring System (NARMS, USA) in 2001 showed that out of 344 *Shigella* isolates available, ~71% of the isolates were found to have developed resistance against 2 or more antibiotics (3). Our central hypothesis is that inhibition of transcription activation of Rns and VirF will prevent the expression of virulence factors in ETEC and *Shigella*, thereby reducing the ability of these pathogens to cause human diseases. Although VirF and Rns have different gene targets, Rns is functionally inter-changeable with VirF and activates the

expression of VirF regulated virulence genes (220), suggesting that inhibitors identified against one protein may inhibit the other. Moreover, sequences of Rns and VirF are very well conserved among different strains of ETEC and *Shigella* respectively. Therefore, identification of inhibitors against Rns and VirF can reduce infections by different other ETEC and *Shigella* strains.

Recently using high-throughput screening the Egan lab identified a compound SE-1 (Fig. 23A) that successfully inhibited AraC family activators RhaS and RhaR. SE-1 specifically inhibited RhaS and RhaR activation *in vivo* in *E. coli* LacZ based reporter gene assays, and blocked DNA binding by both purified full-length RhaS and RhaR proteins (202). In addition, we found that SE-1 inhibited RhaS and the RhaS DNA-binding domain (RhaS-DBD) to the same extent (202). Given that AraC family proteins have a (relatively) conserved DNA binding domain, we hypothesized that SE-1 might exhibit fairly broad inhibition of DNA binding by other AraC family proteins in addition to RhaS and RhaR. Here, I have tested whether this hypothesis applies to ETEC and *Shigella* major virulence activators Rns and VirF.

SE-1 inhibited Rns and VirF activation of a *virB-lacZ* fusion in *E. coli*. SE-1 inhibition of transcription activation by Rns and VirF was first tested using cell-based assays in *E. coli*. These assays were performed by a former graduate student Jeff Skredenske from my lab. To test the SE-1 inhibition of VirF and Rns, β-galactosidase assays were performed for bacteria grown in the absence or presence of inhibitor. Non-pathogenic strains of *E. coli* that carried a single copy of the *lacZ* reporter gene under the control of the *virB* promoter, *virB-lacZ*, in the chromosome (Fig. 23B, Top) and plasmids with IPTG-inducible expression of either Rns or VirF were used. Rns and VirF activation levels were tested from the same *virB* promoter region as both Rns and VirF are functionally interchangeable (220). A control strain carried plasmid-encoded LacI and the *lacZ* reporter gene under the control of a synthetic, LacI-repressible *hts* promoter, *hts-lacZ*

A)



B)

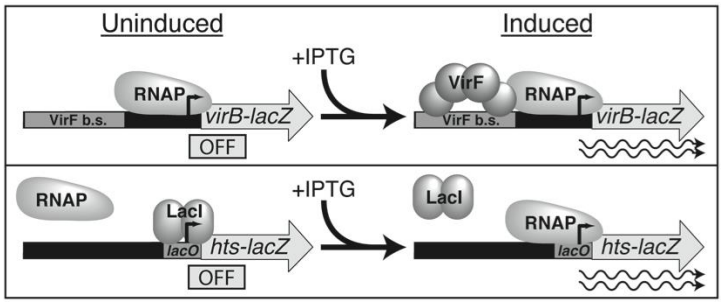


Fig. 23. Reporter fusions used for *in vivo* assays and structure of SE-1. (A). Chemical structure of SE-1, 1-butyl-4-nitromethyl-3-quinolin-2-yl-4*H*-quinoline. (B). VirF-activated *virB-lacZ* fusion (Top) and LacI-repressed *hts-lacZ* control fusion (Bottom), each shown in their uninduced (-) IPTG state (left) and their induced (+) IPTG state (right). Gray rectangles: VirF and LacI (*lacO*) binding sites. Gray arrows: *lacZ* gene expressed from *virB* or synthetic *hts* promoter. Thick black/gray lines: Promoter region DNA. Right angle black arrows: Transcription start sites. Rounded gray shapes: RNAP, VirF or LacI proteins. Wavy lines: active transcription in the presence of IPTG. (Fig 23B; Courtesy: Dr. Susan Egan)

(202) (Fig. 23B, Bottom). This control strain does not require Rns or VirF for LacZ expression and therefore served to distinguish inhibition of overall transcription or β-galactosidase activity from selective inhibition of transcription activation by Rns and VirF. IPTG was added to the cells (to induce Rns/VirF expression or release LacI from repressing *hts-lacZ*) at the same time as inhibitor. Thus, β-galactosidase was maintained at an uninduced level until inhibitor was added, and it was not necessary to wait for decay of preformed β-galactosidase to detect inhibition. Expression of *lacZ* from uninduced and induced controls was set to 0% and 100%, respectively, thus normalizing the effect of inhibitor on the all the fusions. SE-1 inhibition of transcription activation by either of these virulence activators (Rns and VirF) was predicted to decrease *lacZ* expression from their respective strains carrying *virB-lacZ* fusions with little or no effect on *hts-lacZ* expression. In contrast, a decrease in *lacZ* expression from both *virB-lacZ* and *hts-lacZ* would indicate that the inhibition was not specific for Rns and VirF.

We found that SE-1 showed a substantially greater dose-dependent inhibition of the VirF-activated *lacZ* fusion than the LacI-repressed fusion (Fig. 24A). Inhibition of the expression from the VirF-activated fusion had an IC_{50} value of 8 μM. This is a typical potency for a hit from a high throughput screen (low μM to high nM potency range) that has not yet been chemically optimized (221) and more potent than the IC_{50} value of 30 μM found in similar assays with RhaS (202). Similar to the previous report in which we used *hts-lacZ* as a control (202), there was some non-specific inhibition of the LacZ expression from the LacI-repressed fusion at higher doses of SE-1.

Similar to dose-dependent inhibition of the VirF-activated *virB-lacZ* fusion, we found dose-dependent inhibition of the LacZ expression from Rns-activated *virB-lacZ* fusion. The inhibition had an IC_{50} value of 9 μM, suggesting that SE-1 inhibits Rns and VirF to the same

extent *in vivo* in *E. coli* (Fig. 24B). Overall, our cell-based assays suggested that SE-1 inhibits transcription activation by Rns and VirF, at least in *E. coli*, with reasonable selectivity.

SE-1 did not impact *E. coli* growth. Bacterial growth assays were performed to ensure that the observed inhibition of Rns and VirF activation could not be attributed to toxicity toward the strains of *E. coli* used in the cell-based reporter assays. Strains carrying the Rns-activated, VirF-activated and LacI-repressed fusions were grown for 8 h at 37^{o}C in the same minimal medium used for the *in vivo* dose-response assays. We found no detectable effect of SE-1 on growth of any of the strains, indicating that the inhibition observed in the whole cell assays could not be attributed to effects on cell growth (Fig. 25 and Fig. 26). As a control, we tested whether 0.3% DMSO (the solvent for SE-1) had any impact on the growth rate of these strains of *E. coli*, and found that it did not have any impact on the growth rate (Fig. 25 and Fig. 26). The finding that the tested concentrations of SE-1 had no detectable effects on the bacterial growth rates further supports the conclusion that SE-1 exhibits selectivity for Rns and VirF inhibition.

SE-1 inhibited *in vitro* DNA binding by VirF. Electrophoretic mobility shift assays (EMSAs) were performed to investigate whether, similar to RhaS and RhaR (202), SE-1 inhibits DNA binding by purified VirF protein (MBP-VirF). The VirF protein was purified using amylose affinity chromatography (done by Bria Kettle) (Fig. 27), and the DNA tested included the VirF binding site sequence from the *virB* promoter region. Our results indicate that SE-1 was able to fully inhibit DNA binding by VirF in a dose-dependent manner (Fig. 28). The concentration of SE-1 required for half-maximal inhibition of DNA binding was higher than the IC_{50} for the cell-based assays. However, it was not possible to calculate an accurate IC_{50} value from the DNA binding assays due to limitations of the detection method sensitivity, solubility of SE-1, and residual aggregation of the purified VirF protein. We previously showed that SE-1 did not inhibit

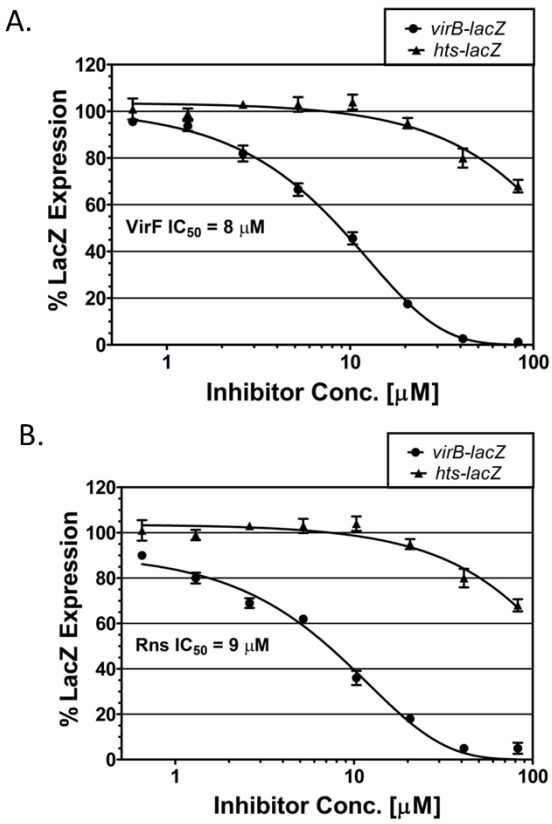


Fig. 24. SE-1 inhibition of *in vivo* Rns and VirF activation in *E. coli*. (A) β-galactosidase (LacZ) activity was assayed at the indicated concentrations of inhibitor SE-1 from two reporter fusions in *E. coli*: VirF-activated *virB-lacZ* (SME4382, circles) and LacI-repressed *hts-lacZ* (SME3359, triangles). (B) β-galactosidase (LacZ) activity was assayed at the indicated concentrations of inhibitor SE-1 from two reporter fusions in *E. coli*: Rns-activated *virB-lacZ* (SME4382, circles) and LacI-repressed *hts-lacZ* (SME3359, triangles). Rns, VirF and LacI were expressed from plasmid pHG165. Activity in the absence of SE-1 was set to 100% in each case, and corresponded to approximately 950 Miller Units for VirF-activated *virB-lacZ,* 1000 Miller Units for Rns-activated *virB-lacZ* and 350 Miller Units for *hts-lacZ.* Results are the average of three independent experiments with two replicates each (Figure courtesy of Dr. Jeff Skredenske).

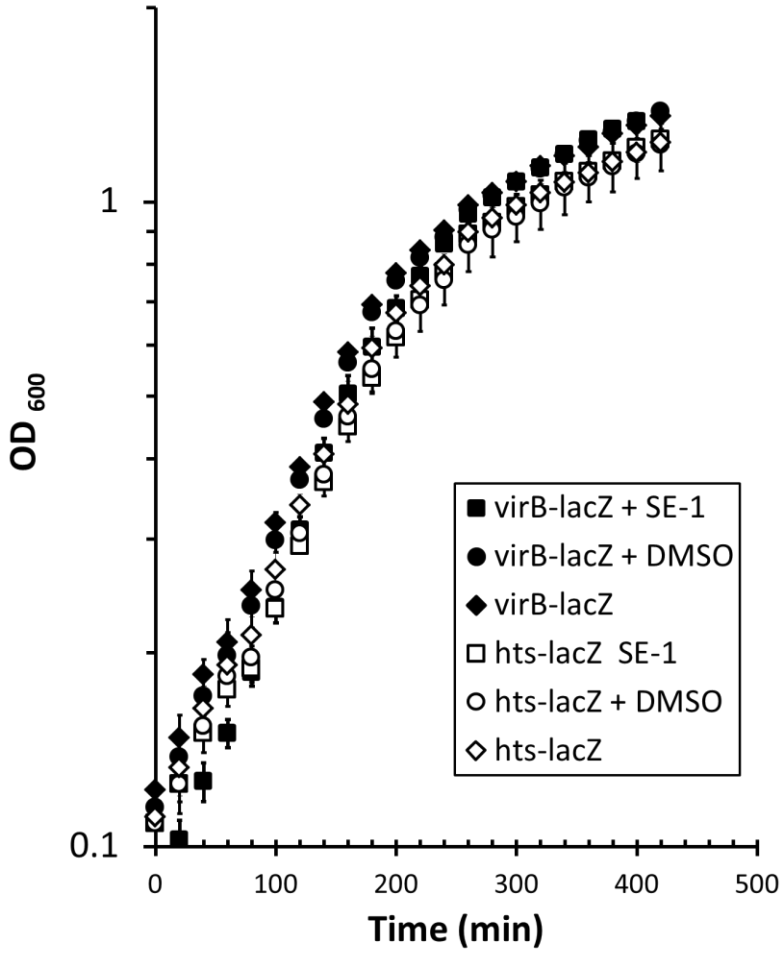


Fig. 25. Growth curves of *E. coli* strains carrying *virB-lacZ* and VirF expressed from pHG165 (SME4382, squares) or *hts-lacZ* and LacI expressed from pHG165 (SME3359, circles) grown in the absence (-) or presence (+) of SE-1 (44 μM), or the absence or presence of DMSO (0.3%). Results are the average of three replicates and error bars represent the standard error of the mean.

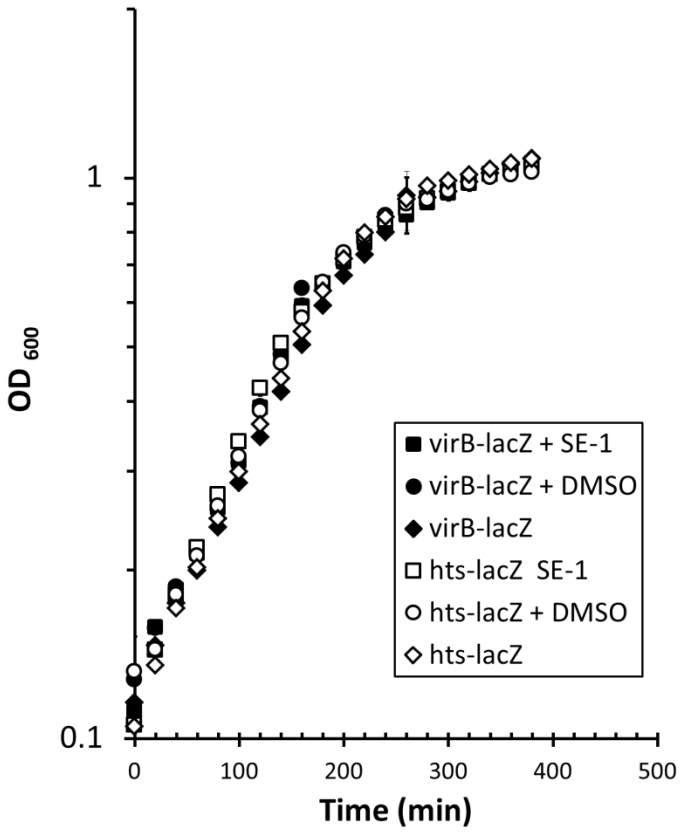


Fig. 26. Growth curves of *E. coli* strains carrying *virB-lacZ* and Rns expressed from pHG165 (SME4383, squares) or *hts-lacZ* and LacI expressed from pHG165 (SME3359, circles) grown in the absence (-) or presence (+) of SE-1 (44 μM), or the absence or presence of DMSO (0.3%). Results are the average of three replicates and error bars represent the standard error of the mean.

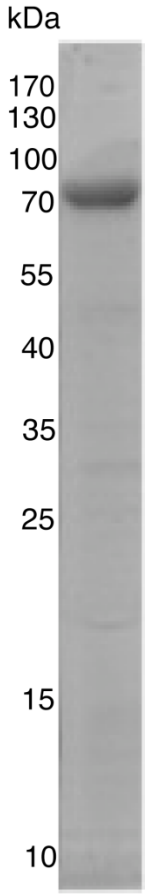


Fig. 27. MBP-VirF protein (right lane) after purification using amylose affinity chromatography column. Pre-stained protein ladder (left lane, Thermo Fisher Scientific) with molecular weights of 170, 130, 100, 70, 55, 40, 35, 25, 15 and 10 kDa (from top). MBP-VirF protein has a predicted molecular mass of 73 kDa. (Figure courtesy of Bria Kettle)

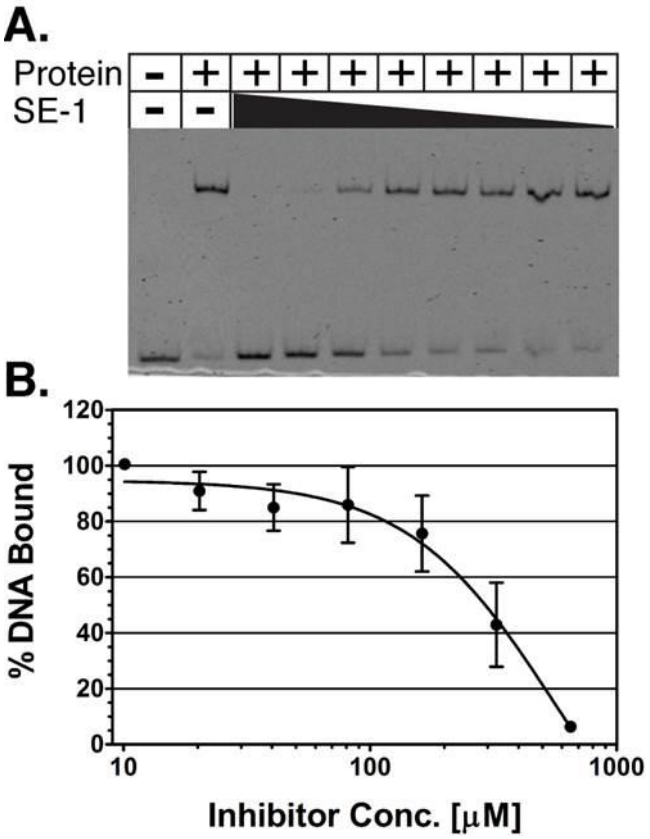


Fig. 28. SE-1 inhibition of in vitro DNA binding by VirF. (A) A representative EMSA gel picture. Black triangle represents decreasing concentrations of SE-1, from 1.3 mM to 10 μM, with serial two-fold dilutions. (B) Binding of purified VirF to a DNA fragment containing the VirF binding site from the *virB* promoter was assayed using electrophoretic mobility shift assays (EMSAs) with inhibitor SE-1 concentrations from 10 to 650 μM (serial 2-fold dilutions). DNA and protein were at final concentrations of 2 and 300 nM, respectively. The DNA shifted (bound by VirF) was quantified, and the value at the lowest concentration of SE-1 was set to 100%. Results are the average of three independent replicates.

DNA binding by the LacI or CRP proteins (202). LacI and CRP are not members of the AraC family, and each are founding members of their own protein families (222-224). Overall, our EMSA results indicate that SE-1 can block the ability of VirF to bind to its specific DNA site.

SE-1 inhibited *in vitro* DNA binding by Rns. Electrophoretic mobility shift assays (EMSAs) were performed to investigate whether SE-1 inhibits DNA binding by purified MBP-Rns protein. MBP-Rns was purified using amylose affinity column as described above. Specific labeled DNA template (one IR labeled and one non-IR labeled oligo) containing Rns binding site from *cfaA* promoter region was used. This promoter is a native target for Rns activation (unlike the *virB* promoter tested in the *in vivo* assays) where Rns binds and activates the transcription of genes encoding CFA/I fimbriae that mediate the host cell attachment to intestinal epithelium (225, 226). EMSA results indicated that SE-1 inhibits DNA binding by Rns in a dose-dependent manner (Fig. 29). The IC_{50} value of inhibition was found to be similar to half maximal inhibition of DNA binding by VirF in EMSAs.

Molecular docking prediction of SE-1 binding site. We used BSP-SLIM (227) to predict where SE-1 may bind in the I-TASSER predicted structures of Rns and VirF. Given that, SE-1 inhibits the activity of the DBD to a similar extent as full-length protein (based on experiments on RhaS) (202), we focused only on the Rns-DBD and VirF-DBD in our docking simulations. SE-1 identified was from pub chem library and ZINC library (ZINC library from University of California, SanFransisco, CA) with codes STK330322 and 4046855, respectively. The top models generated by BSP-SLIM predicted SE-1 to bind in a pocket between the two helix-turn-helix motifs (HTH). Figure 30 and 31 shows the binding of SE-1 in the predicted structures of Rns and VirF, respectively. Binding of SE-1 between two HTH motifs could sterically prohibit the binding of Rns-DBD and VirF-DBD to their respective promoter DNA. In the past, docking

predictions of SE-1 carried out by Jeff Skredenske using SwissDock program (www.swissdock.ch, Molecular Modeling Group, Swiss Institute of Bioinformatics, Lausanne, Switzerland) on other AraC family proteins including I-TASSER predicted full-length RhaS and RhaS-DBD structures, AraC-DBD NMR structure (PDB 2K9S) and MarA co-crystal structure with DNA (PDB 1BL0) placed SE-1 between two HTH motif. All these results suggest that SE-1 could potentially interfere with DNA binding of AraC family proteins including Rns and VirF.

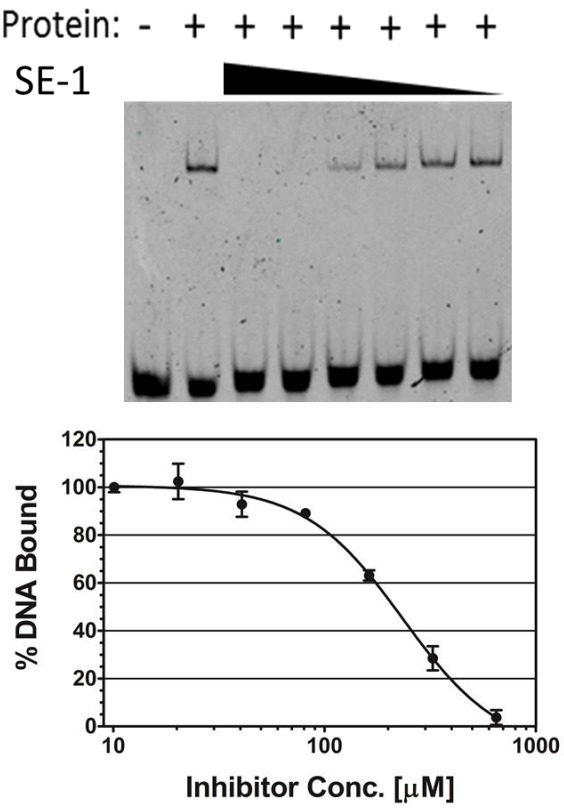


Fig. 29. SE-1 inhibition of in vitro DNA binding by Rns. (A) A representative EMSA gel figure. Decreasing concentrations of SE-1, from 1.3 mM to 10 μM, with serial two-fold dilutions were represented by black triangles. (B) Binding of purified Rns to a labeled DNA fragment containing the Rns binding site from the *cfaAp* was assayed using EMSAs with 10 to 650 μM (serial 2-fold dilutions) inhibitor concentrations. DNA and protein were at final concentrations of 2 and 300 nM, respectively. The Rns bound DNA (Shifted band) was quantified. The %DNA binding at the lowest concentration of SE-1 was set to 100%. Results are the average of three independent experiments.

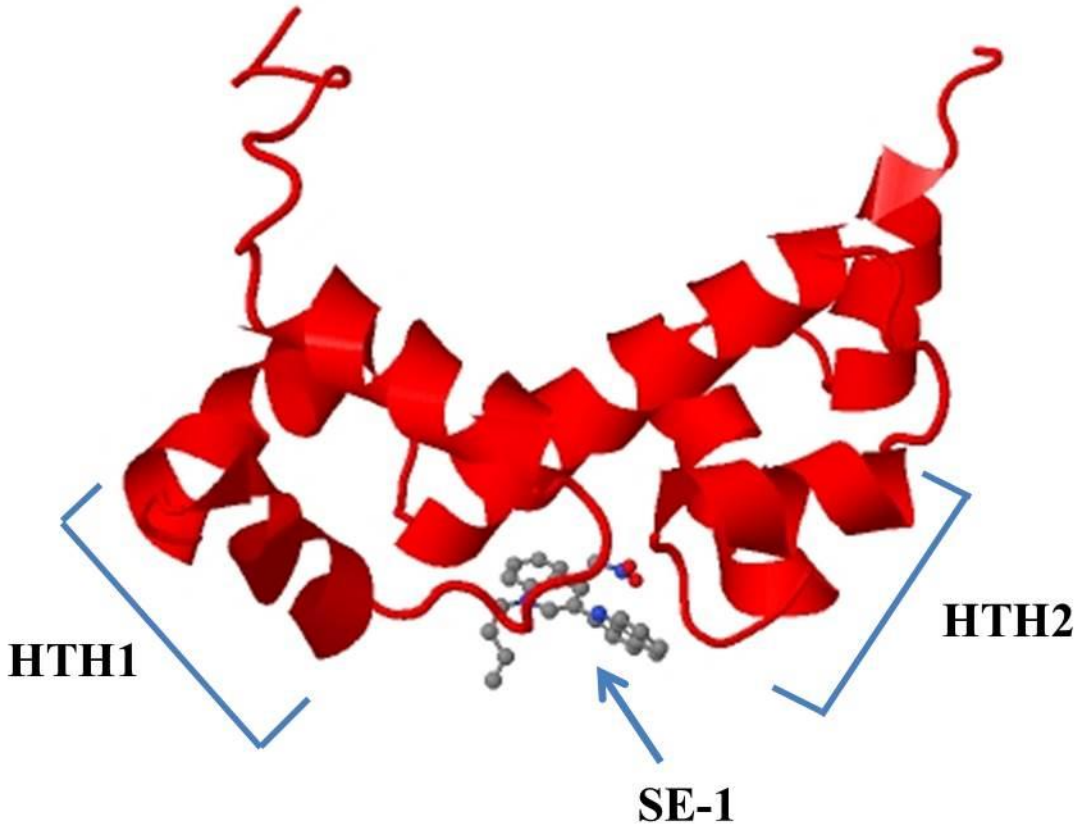


Fig. 30. Prediction of SE-1 binding site on Rns-DBD. A computational docking program BSP-SLIM was used to predict the binding site of SE-1 on I-TASSER (157) predicted structure of Rns-DBD. Rns-DBD was labeled in red color and SE-1 was labeled in grey color.

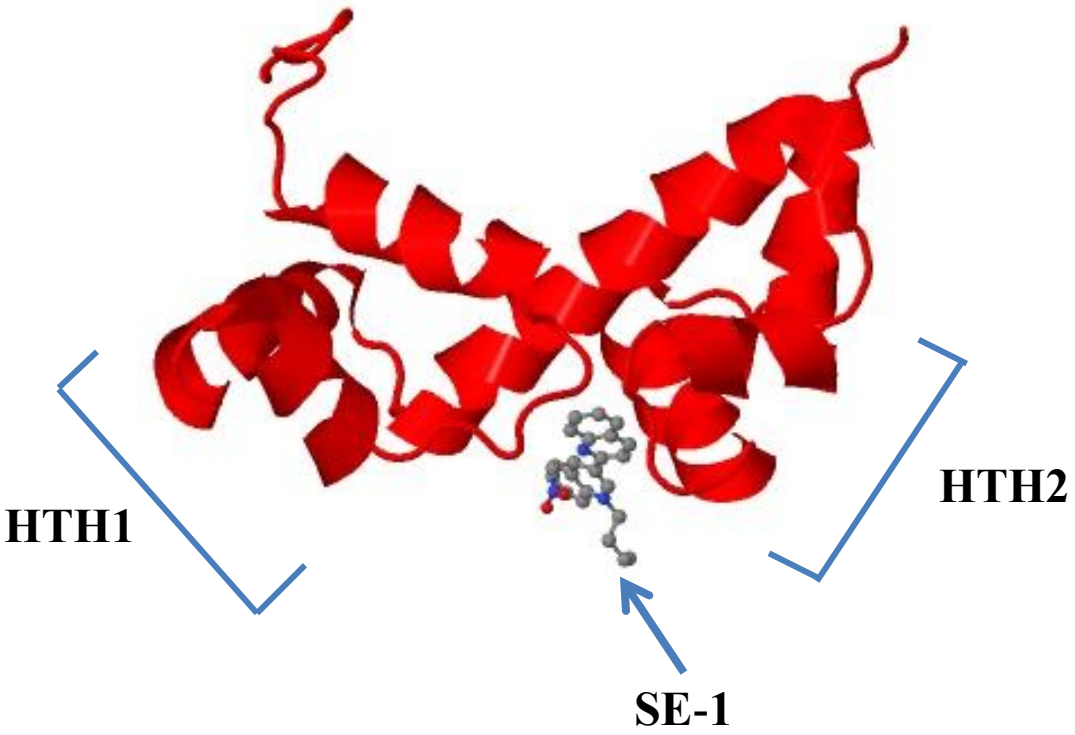


Fig. 31. Prediction of SE-1 binding site on VirF-DBD. A computational docking program BSP-SLIM was used to predict binding site of SE-1 on I-TASSER (157) predicted structure of VirF-DBD. VirF-DBD was labeled in red color and SE-1 was labeled in grey color.

SE-1 inhibited VirF activation of a *virB-lacZ* fusion in *Shigella*. Our cell-based assays in *E. coli* indicated that SE-1 was able to effectively inhibit activation by VirF at the *virB* promoter region. To test whether a similar inhibition occurred in *Shigella*, we assayed the effect of SE-1 on VirF activity in a *S. flexneri* strain carrying a *virB-lacZ* fusion. β-galactosidase assays were performed on bacterial cultures grown at 37^{o}C in the absence or presence of varying concentrations of inhibitor. Our results show a dose-dependent inhibition of *virB-lacZ* expression, with a maximum inhibition of greater than two fold at 40 μM SE-1 (Fig 32). This inhibition was most likely due to inhibition of VirF activity. We calculated an IC_{50} value of approximately 30 μM, which is somewhat higher than the 8 μM IC_{50} achieved in *E. coli*. *S. flexneri* samples grown in 80 μM SE-1 were not analyzed as they exhibited growth defects. Cultures of *Shigella* grown at 30^{o}C in the absence of SE-1 were used to identify the basal level of *virB-lacZ* expression (161, 162, 166). These assays, similar to the cell-based assays in *E. coli*, indicate that SE-1 inhibits transcription activation by VirF at the *virB-lacZ* promoter region, and further provide evidence that SE-1 is effective in *Shigella*.

SE-1 reduced VirF-regulated virulence gene expression in *Shigella*. Our next goal was to test the impact of SE-1 on expression of other VirF-regulated genes in *Shigella*. Previous reports have shown that *icsA* and *virB* genes are direct targets of VirF activation at 37^{o}C (161). VirB, in turn, activates the expression of many operons that play a crucial role in *Shigella* pathogenesis (148). Thus, any compound that inhibits transcription activation by VirF should also affect the expression levels of genes that either directly or indirectly require VirF for their expression. To test this hypothesis, we used qRT-PCR to quantify mRNA levels of two direct VirF target genes, *icsA* and *virB,* and two indirect VirF target genes, *icsB* and *ipaB* (both directly activated by VirB) (148).

Before testing the effect of SE-1 on the expression of VirF regulated genes, we first wanted to ensure that the expression of VirF regulated genes were indeed higher at 37^{o}C relative to 30^{o}C. In addition, we wanted to identify the growth stage at which the relative expression of VirF regulated genes reaches to maximal levels. Therefore, we performed real-time PCR analysis of *virB* and *icsB* expression levels from *S. flexneri ipgD*- grown until OD_{600} of 0.55, 0.75 and 1.0 at both 30^{o}C and 37^{o}C. Quantification of mRNA levels showed that *virB* and *icsB* genes were expressed at higher levels at 37^{o}C compared to 30^{o}C (Table 8). In addition, expression of *virB* and *icsB* was also higher in samples grown until late log phase (OD_{600} ~1.0) (Table 8).

To test the effect of SE-1 on VirF regulated genes, *S. flexneri* samples were grown to an OD_{600} of ~1.0 in the absence or presence of two different concentrations of SE-1 (20 μM, 40 μM) and the mRNA levels from each of these genes were quantified. *S. flexneri* samples grown at 30^{o}C in the absence of SE-1 were included to illustrate the basal expression levels of expression of these genes. Results were normalized to the *gapA* (encodes glyceraldehyde-3-phosphate dehydrogenase) and *rrsA* (encodes 16S rRNA) genes, which are two constitutively expressed genes that are commonly used as qRT-PCR controls in *Shigella* (211-213). Example data and calculated values are shown in Table 9.

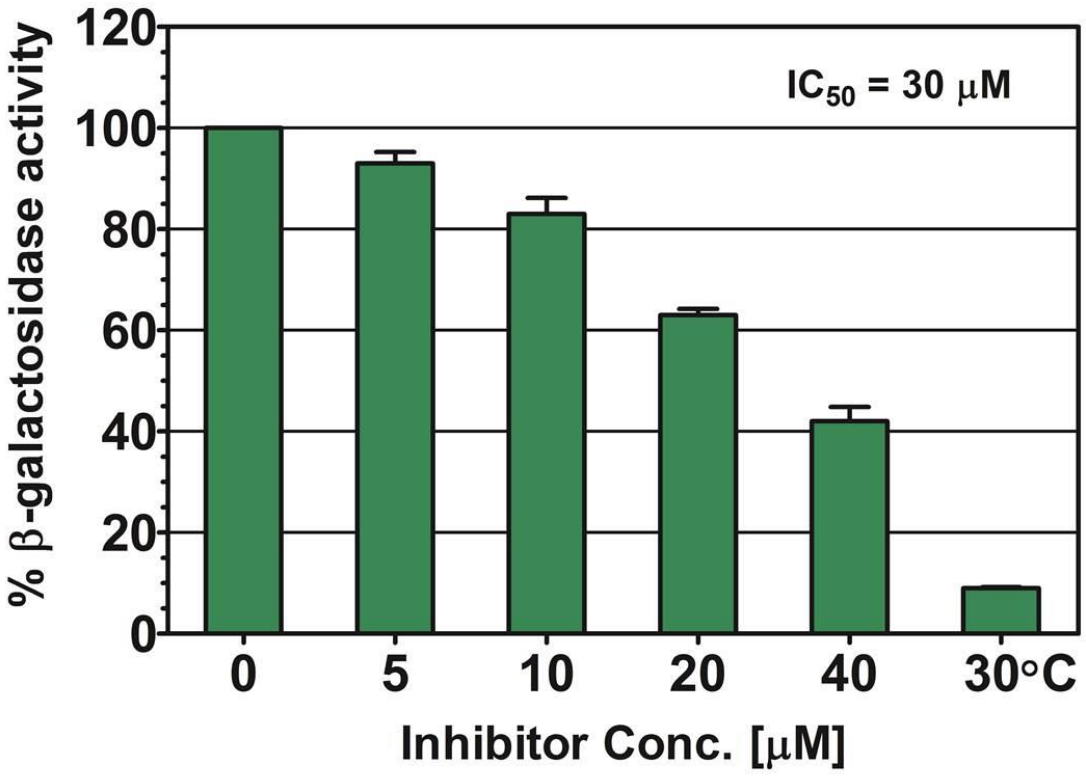


Fig. 32. SE-1 inhibition of *in vivo* VirF activation in *Shigella*. β-galactosidase (LacZ) activity was assayed at the indicated concentrations of inhibitor SE-1 from a *virB-lacZ* transcriptional fusion in *Shigella*, with native VirF expression. Activity in the absence of SE-1 was set to 100% and corresponded to approximately 3,000 Miller Units. *Shigella* grown at 30^{o}C illustrates basal expression of *virB-lacZ*. Results are the average of three independent experiments with two replicates each.

Table 8. Real-time PCR analysis of expression of *virB* and *icsB* genes at 37°C relative to 30°C.

	Expression at 37°C relative to 30°C at different OD_{600}		
Gene	OD_{600} =0.55	OD_{600}= 0.75	OD_{600}=1
virB	0.84	6.92	9
icsB	4	26.2	33

S. flexneri samples grown until OD_{600} of approximately 0.55, 0.75 and 1 were compared for the expression of *virB* and *icsB* genes at 37°C relative to 30°C. Data was normalized to *gapA* internal control. Level of expression at 30°C was set to 1 and compared to expression at 37°C.

Table 9. Analysis of example real-time PCR gene expression data.

Gene	SE-1 (μM)	Ct	ΔΔCt[1]	ΔΔCt[2]	AverageΔΔCt	RQ ($2^{-\Delta\Delta Ct}$)
gapA	0	26.70		0	0	1
gapA	20	27.33		1	1	0.5
gapA	40	26.76		-0.69	-0.69	0.62
rrsA	0	18.91	0		0	1
rrsA	20	18.55	-1.0		-1.0	2
rrsA	40	18.29	0.69		0.69	1.62
virB	0	22.00	0	0	0	1
virB	20	23.26	0.63	1.63	1.13	0.46
virB	40	23.18	1.12	1.81	1.47	0.36
icsA	0	27.70	0	0	0	1
icsA	20	27.88	-0.46	0.54	0.04	0.97
icsA	40	28.59	0.82	1.52	1.17	0.44
icsB	0	26.17	0	0	0	1
icsB	20	26.25	-0.56	0.44	-0.06	1.04
icsB	40	27.04	0.80	1.49	1.14	0.45
ipaB	0	24.32	0	0	0	1
ipaB	20	25.57	0.62	1.61	1.11	0.46
ipaB	40	26.23	1.85	2.54	2.20	0.22

[1]ΔΔCt values after normalizing data to *gapA*

[2] ΔΔCt values after normalizing data to *rrsA*

ΔΔCt = (Ct,Target - Ct,Internal Control) SE-1 - (Ct,Target - Ct, Internal Control) no SE-1.

The values in the table are for a single replicate. The RQ values from three such replicates were averaged to get the average gene expression levels and to calculate the standard error.

Our qRT-PCR results showed that the *icsA* expression level remained essentially unchanged at 20 μM inhibitor, but was significantly reduced ($P < 0.05$) at 40 μM inhibitor (Fig. 33A). Expression of *virB*, another direct target of VirF exhibited a significant decrease in expression with increasing concentrations of inhibitor, and a maximal inhibition of more than two fold at 40 μM SE-1 ($P < 0.05$) (Fig. 33B)**.** Expression of *icsB*, an indirect target of VirF (183, 184), showed a significant two-fold reduction at 40 μM SE-1 ($P < 0.05$) (Fig. 34A). The greatest inhibition was detected for expression of *ipaB*, another indirect target of VirF, which was reduced by up to 4-fold at 40 μM inhibitor ($P < 0.05$) (Fig. 34B). As a control, we normalized each of the internal control genes relative to the other internal control gene. We found that there was no decrease in the expression of either control gene with 40 μM SE-1 (Fig. 35), arguing that SE-1 did not result in a global decrease in gene expression. Together, our data demonstrate that SE-1 is capable of reducing the expression of at least four VirF-regulated virulence-associated genes in *Shigella*, presumably through its inhibition of transcription activation by VirF.

SE-1 inhibited host cell invasion by *Shigella*. Our qRT-PCR experiments with SE-1 showed a decrease in expression of virulence genes in *Shigella*. In order to test the impact of this decrease in virulence gene expression on the ability of *Shigella* to invade host cells, invasion assays (228) were performed using the mouse fibroblast cell line L-929. These experiments were performed in collaboration with Dr. Scott Hefty's lab and a graduate student Ichie Osaka performed these experiments. L-929 cells were used to study the invasion of *Shigella* into the host cells (229-232). The *S. flexneri* strain used in the assays was an *ipgD*$^{-}$ strain, which has hemolysis and invasion properties that are similar to wild type, but is safer to work with due to its reduced ability to cause human infection (233-235). The cells were grown at 30^{o}C overnight

and then diluted and grown at 37^{o}C (to induce expression of VirF-regulated genes) in the presence or absence of various concentrations of SE-1. In addition, *S. flexneri* strain SB116 (*mxiH*$^{-}$) was used as a negative control for invasion since it does not form a T3SS needle and thereby is unable to invade host cells (46). *S. flexneri* grown at 30^{o}C was used as a second negative control for invasion, as these bacteria have reduced expression of VirF and VirB and thereby are defective for invasion of host cells (148). Our results showed a dose-dependent decrease in invasion of *Shigella* into L-929 cells upon addition of SE-1 (Fig. 36). SE-1 resulted in invasion decreases of 1.7-fold at 20 μM and 3-fold at 40 μM relative to no inhibitor. As expected, both of the negative controls (wild type *S. flexneri* grown at 30^{o}C and the *mxiH*$^{-}$ strain) showed substantially decreased invasion relative to the wild type strain grown at 37^{o}C. Similar to our growth assays in *E. coli*, we found that 0.3% DMSO did not detectably slow the growth of *Shigella* (data not shown), and SE-1 resulted in no detectable decrease in the *Shigella* growth rate at concentrations up to 40 μM (Fig. 37).

Overall, our results support the hypothesis that SE-1 inhibits *Shigella* invasion by decreasing the expression of the genes in the VirF regulon, including the genes that encode the T3SS. Thus, *Shigella* grown under inducing conditions (37^{o}C) without SE-1 would not be inhibited by addition of SE-1 unless growth with SE-1 was continued until the loss of preformed T3SS. To test this hypothesis, *Shigella* were grown at 37^{o}C in the absence of inhibitor, and then *Shigella* and SE-1 were added at the same time to L-929 cells. This control also tested whether SE-1 had any impact on the L-929 cells that affected invasion by *Shigella*. Our result suggest that the inhibition of invasion by SE-1 was likely not due to post-transcriptional effects on *Shigella* or effects on the eukaryotic cells, as this control sample invaded to the same extent as *Shigella* grown in the absence of inhibitor (Fig. 36). Overall, our results support the hypothesis

that SE-1 decreases the expression of VirF-regulated virulence genes, and that this decrease in turn resulted in a reduction in the ability of *Shigella* to invade host cells.

In principle, the decrease in L-929 cell invasion by *Shigella* might be the result of SE-1 compromising host cellular processes, and thereby decreasing the viability of the host cells. To test this possibility, the effect of SE-1 on the viability of L-929 cells was assayed using an AlamarBlue® cell viability assay (Invitrogen, Grand Island, NY). The AlamarBlue® assay is a fluorescence-based assay of the viability of host cells based on their metabolic activity (236). During invasion assays, the L-929 cells were only exposed to SE-1 concentrations of 0.2 and 0.4 μM, due to dilution of the SE-1-containing *Shigella* culture into the invasion assays. AlamarBlue® assays indicated that these concentrations of SE-1 did not detectably decrease the viability of L-929 cells (Fig. 38). In fact, a concentration 100-fold higher than this (40 μM) had no effect on the host cell viability. These results support the hypothesis that the effect of SE-1 on invasion by *Shigella* was not an indirect effect on L-929 cell viability. Further, the results provide evidence that SE-1 did not have detectable toxicity toward the L-929 cells at concentrations up to 40 μM.

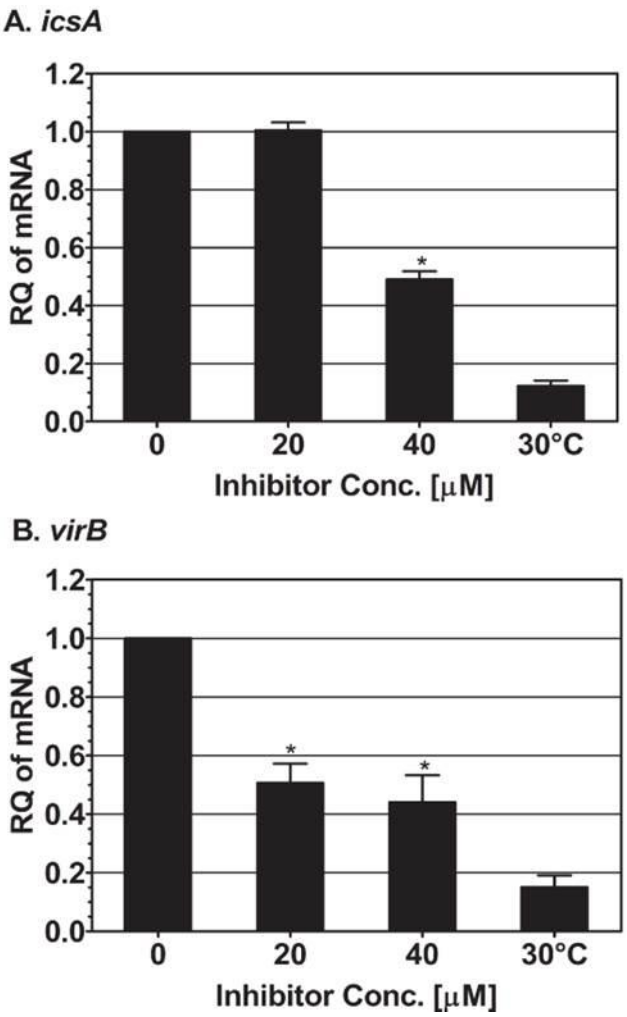


Fig. 33. SE-1 inhibition of VirF regulated virulence gene expression in *Shigella*. qRT-PCR was used to determine the relative quantity (RQ) of expression of *icsA* (A), *virB* (B), in *Shigella* grown at 30°C (without SE-1) and at 37°C (without or with 20 and 40 μM SE-1). RNA levels were normalized to two constitutively expressed genes, *gapA* and *rrsA*. RNA levels were set to 1 in the absence of SE-1. *Shigella* grown at 30°C illustrates basal expression of each gene. Error bars represent the standard error of the mean calculated from three independent replicates. Significance of inhibitor-treated values was calculated using a non-parametric Mann-Whitney test, with P-values less than 0.05 considered significant (*). Results are the average of three independent experiments with one replicate each.

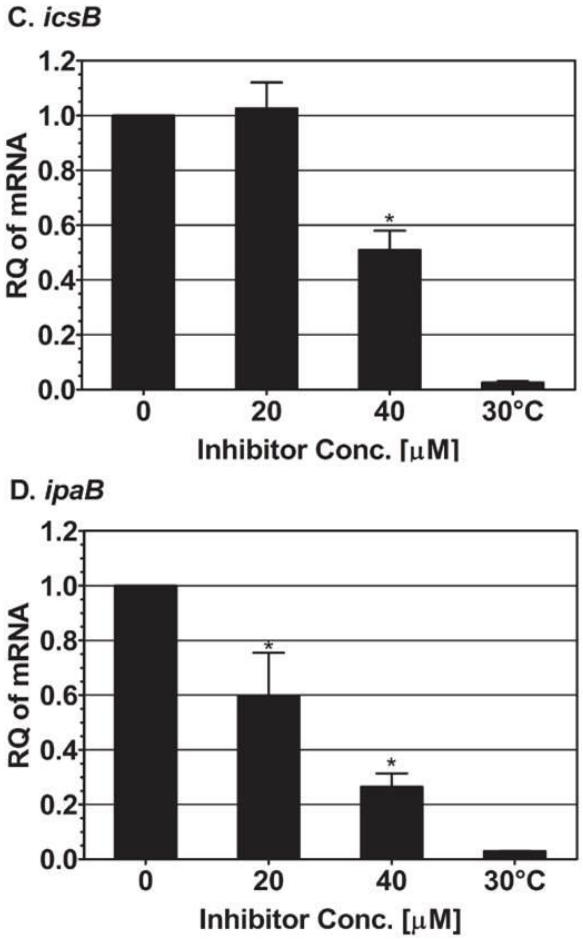


Fig. 34. SE-1 inhibition of VirF regulated virulence gene expression in *Shigella*. qRT-PCR was used to determine the relative quantity (RQ) of expression of *icsB* (A), *ipaB* (B), in *Shigella* grown at 30°C (without SE-1) and at 37°C (without or with 20 and 40 μM SE-1). RNA levels were normalized to two constitutively expressed genes, *gapA* and *rrsA*. RNA levels were set to 1 in the absence of SE-1. *Shigella* grown at 30°C illustrates basal expression of each gene. Error bars represent the standard error of the mean calculated from three independent replicates. Significance of inhibitor-treated values was calculated using a non-parametric Mann-Whitney test, with P-values less than 0.05 considered significant (*). Results are the average of three independent experiments with one replicate each.

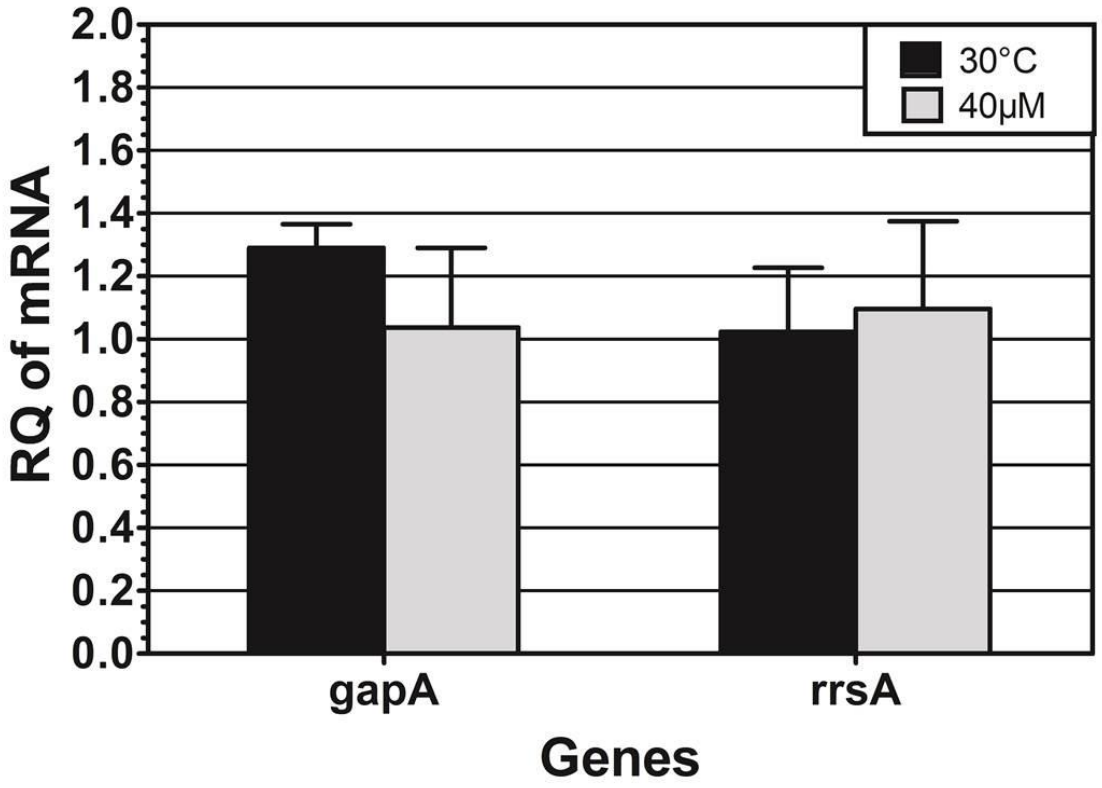


Fig. 35. Relative quantification of mRNA levels of the constitutively expressed genes *gapA* and *rrsA* of *Shigella*, with *gapA* normalized to the *rrsA* values, and *vice versa*. Cultures were grown at 30^{o}C (Black bars) or at 37^{o}C with 40 µM SE-1 (Grey bars). Values are relative to the mRNA levels of *gapA* (for *gapA*) or *rrsA* (for *rrsA*) for *Shigella* grown at 37^{o}C with no SE-1, which were set to one. Error bars represent the standard error of the mean calculated from three independent replicates.

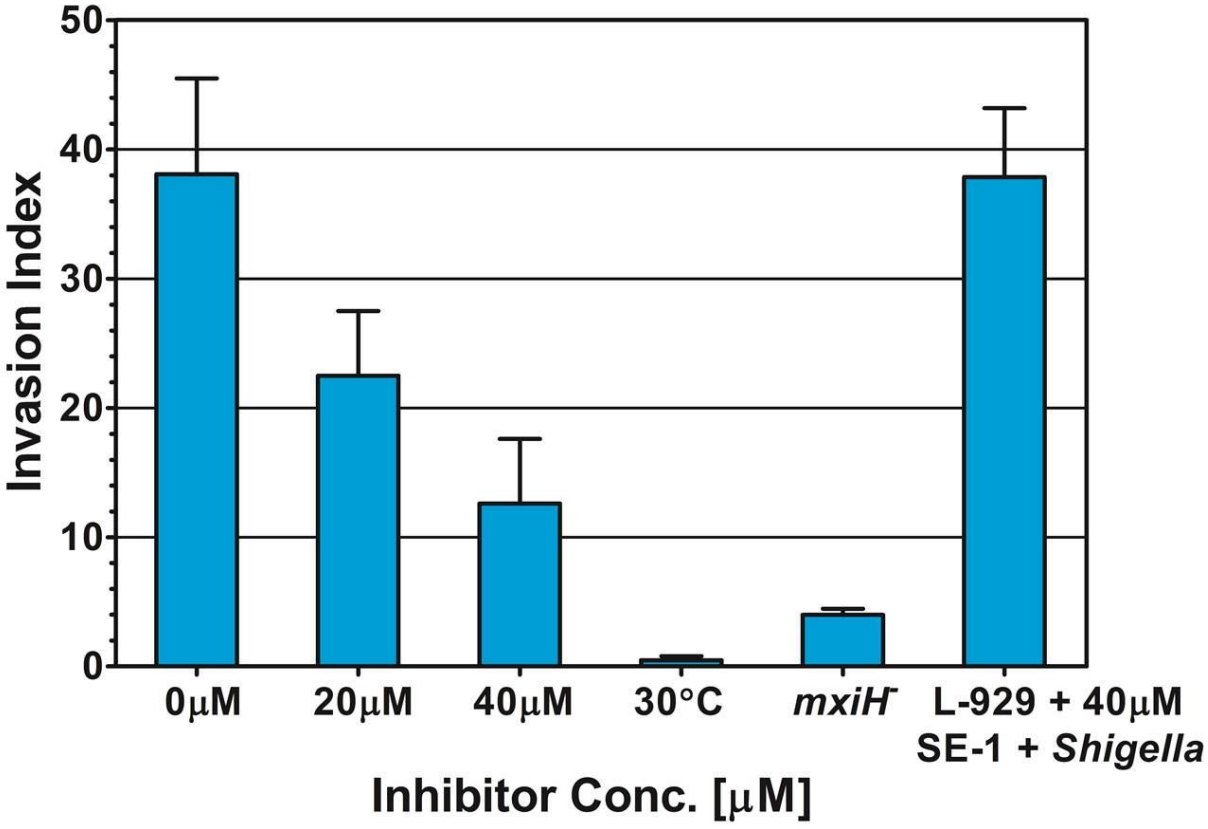


Fig. 36. SE-1 inhibition *of Shigella* host cell invasion. The invasion index of *Shigella* into L-929 (mouse fibroblast) cells was determined by gentamycin protection assay. Cultures were grown in the absence or presence of SE-1 (20 μM and 40 μM). A *mxiH*⁻ *Shigella* strain (SH116) and an *ipgD*⁻ strain (SME4331) grown at 30°C were used as negative controls of invasion. As an additional control, 40 μM SE-1 and *Shigella* grown at 37°C without inhibitor were added at the same time to L-929 cells to test the effect of inhibitor on invasion by *Shigella* with a preformed T3SS and on host cells. Assays in were performed in triplicate and error bars represent the standard error of the mean. (Figure courtesy of Dr. Ichie Osaka)

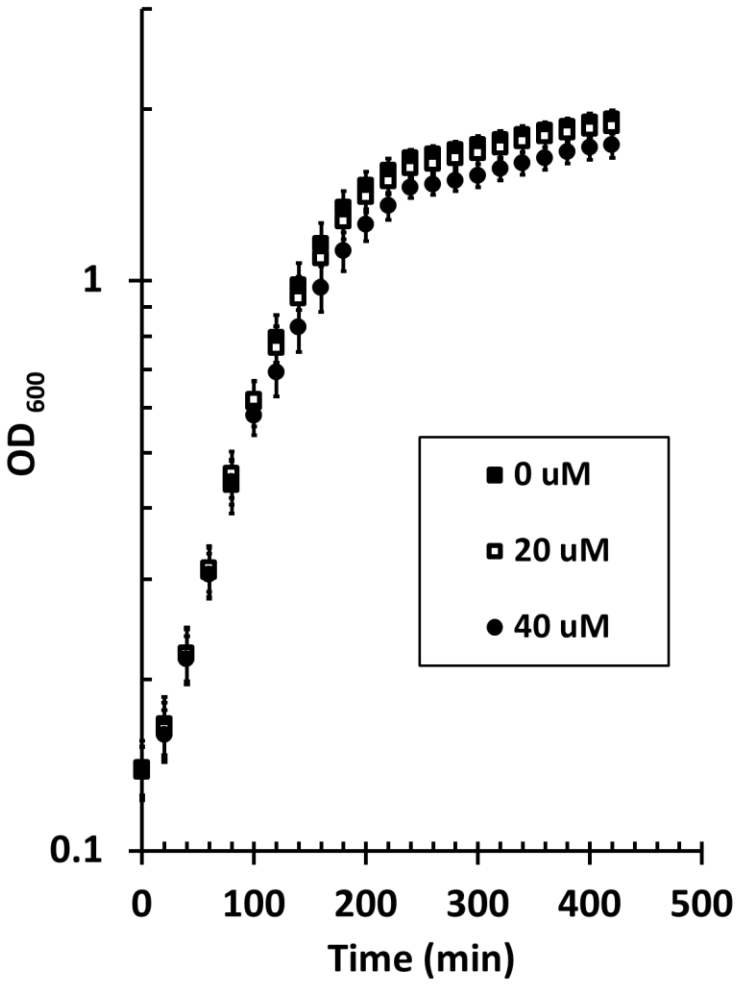


Fig. 37. Growth curves of *Shigella* strain carrying *ipgD*$^{-}$ (SME4331) in the absence (filled square) or presence of SE-1 (20 μM, open squares; 40 μM, filled circles). Results are the average of three replicates and error bars represent the standard error of the mean.

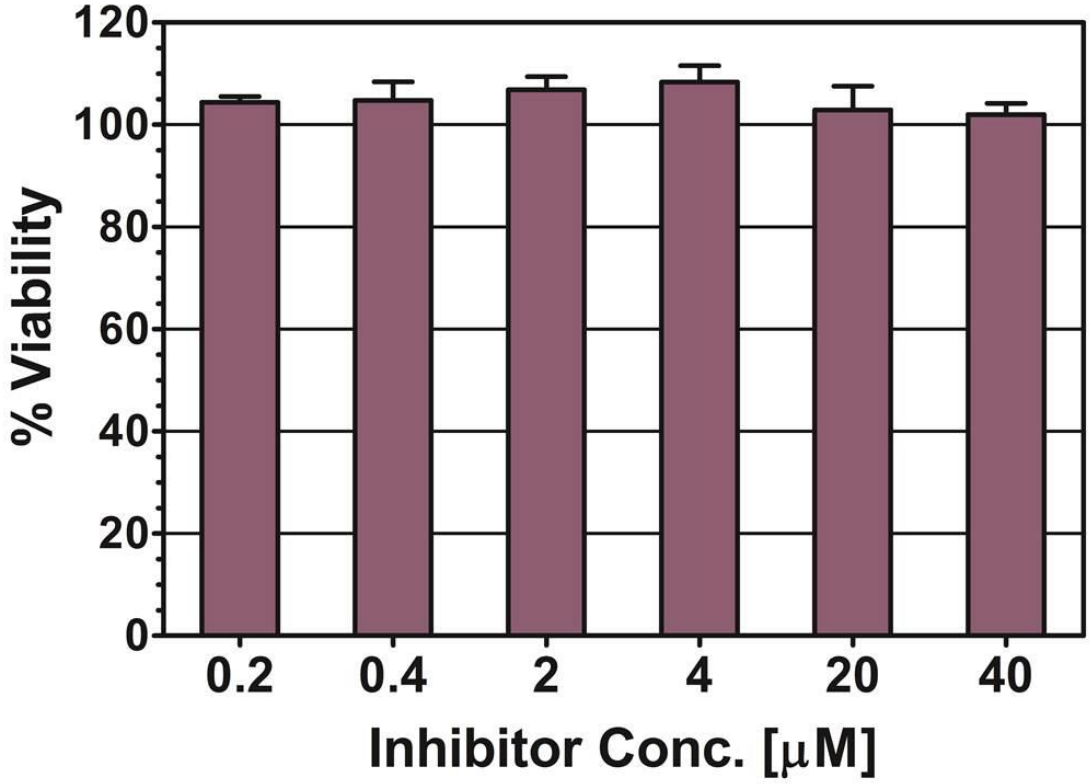


Fig. 38. SE-1 does not inhibit host cell metabolism. AlamarBlue® assays to test the effect of SE-1 (at the indicated concentrations) on the metabolic activity (viability) of L-929 cells. The absorbance level in the absence of inhibitor and the presence of DMSO was set to 100%. Assays in were performed in triplicate and error bars represent the standard error of the mean. (Figure courtesy of Dr. Ichie Osaka)

Chapter 6

Functional analysis of RhaR N-terminal domain Arm region

RhaR activates the transcription of rhaSR operon in response to its effector L-rhamnose. Evidence suggests that L-rhamnose binds at the NTD and transcription activation is carried by DBD in the full length RhaR. The separation of L-rhamnose binding and transcription activation activities between the two domains led us to hypothesize that the signal of L-rhamnose binding in the NTD is somehow communicated to the DBD (allosteric signal), and thus causing structural or conformational changes required by the DBD to activate transcription. We earlier discovered that deletion of RhaR extension (first 34 amino acids) does not play any role in transmitting this signal. However, we could not identify the role of RhaR residues that align with the AraC Arm because of defects in protein levels in RhaR variants with deletions in the aligned region (200).

Site-directed random mutagenesis of RhaR Arm region. To investigate the potential role of the proposed RhaR Arm residues (residues 35-45) (Fig. 39) in inter-domain interactions, I performed site-directed random mutagenesis of residues in this region. I randomized one codon at a time in the RhaR Arm and generated multiple variants at each position to understand the role each residue plays in potentially transmitting the allosteric signal. Variants were first isolated in the strain SME2525 [λϕ (*rhaS-lacZ*)Δ128 Δ(*rhaSR*)::Km *recA*::*cat*]. This strain included a CRP binding site upstream of RhaR binding site and has higher activation levels as CRP activates the expression of *rhaSR* operon both in the absence (-) and presence (+) of rhamnose. The higher activation levels helped in identification of variants with different levels of activation on indicator X-gal plates. After identification of variants in SME2525, variants were ultimately assayed for activation levels by β-galactosidase reporter gene assays in strain SME3160 [λϕ (*rhaS-lacZ*)Δ85 Δ(*rhaSR*)::Km *recA*::*cat*] which does not have a CRP binding site. A comparison of activation levels for both of these strains, with activation by wild-type RhaR, is presented in Table 10.

Fig. 39. Alignment of RhaR and AraC N-terminal regions.

Amino acid sequence of N-terminal residues of RhaR and AraC (aligned by I-TASSER program) (157). Arm residues in both RhaR and AraC were labeled in green. Residues before the Arm region in RhaR represents the proposed extension region.

Table. 10. Comparison of RhaR activation of *lacZ* expression *rhaS* promoter regions with and without CRP binding site.

Fusion	(-) rhamnose[a]	(+) rhamnose[a]
(*rhaS-lacZ*)Δ128	166±9	875±90
(*rhaS-lacZ*)Δ85	1.5±0.4	8.5±0.5

[a]β-galactosidase activity (miller units) levels in the absence and presence of rhamnose for samples expressing RhaR from a low copy number plasmid pHG165 at different *rhaS-lacZ* fusions. Miller unit values were presented as mean ± standard error of three independent experiments with two replicates each.

The increase in the RhaR activation levels in the presence of rhamnose could be because of stimulatory inter-domain interactions, likely involving Arm residues, (+) rhamnose or inhibitory inter-domain interactions, likely involving the Arm, (-) rhamnose. Variants at an Arm residue involved in stimulatory inter-domain interactions (+) rhamnose are expected to have wild type activation levels (-) rhamnose and decreased activation levels (+) rhamnose. Variants at an Arm residue involved in inhibitory inter-domain interactions (-) rhamnose are expected to have increased activation levels (-) rhamnose and wild type activation levels (+) rhamnose. Two other possible outcomes from the site-directed random mutagenesis are variants with no defects in activation [wild type activation levels both (-) and (+) rhamnose] and variants with similar defects in activation [decreased activation levels both (-) and (+) rhamnose]. Variants with decreased activation levels both (-) and (+) rhamnose probably have stability defects or decreased protein levels or decreased DNA binding and thereby are not considered important in inter-domain allosteric signaling.

I mutagenized all 11 residues in the proposed RhaR Arm and identified a total of 46 unique variants. Variants with changes in activation levels of 1.5 fold and a non-over lapping standard deviation intervals relative to the wild type were considered to differ significantly. Out of the total 46 variants, only two variants F43C and A44G had significantly decreased activation levels both (-) and (+) rhamnose, suggesting that the Arm region may not be involved in the stability of the RhaR (Table 11). In addition, our results showed that none of the Arm variants had defects in activation only (+) rhamnose, suggesting that the residues in the Arm likely are not involved in making stimulatory inter-domain interactions in the (+) rhamnose state (Table 11).

In 5 out of 11 residues mutagenized, we found a total of nine variants with increased

activation levels (-) rhamnose and wild type activation levels (+) rhamnose [Table 11, Up (-)]. These include four variants at residue L35 and one variant each at residues L37, K39, D41, F42 and F43. At L35, in addition to the four variants with increased activation levels, the other two variants isolated at L35 had small increases in activation levels (-) rhamnose (although the increases were not significant). This indicates that the side chain at position L35 may be involved in making inter-domain interactions (-) rhamnose that lead to inhibition of RhaR activation, and contribute to the low basal activity of RhaR (-) rhamnose. All the other variants with increased activation levels (one each at positions L37, K39, D41, F42 and F43) in RhaR Arm are likely to be gain-of-residue-function variants (Table 11).

An unexpected phenotype was observed in 15 out of the 46 variants in the RhaR Arm region. In these variants, RhaR activation was significantly decreased (-) rhamnose while maintaining wild type activation levels (+) rhamnose [Table 11, Down (-)].The greater defects in activation by RhaR variants in the (-) rhamnose state relative to the (+) rhamnose state could be explained by the differences in the binding affinity of RhaR to its DNA in those two states. In the (-) rhamnose state RhaR binds to DNA with less affinity relative to the (+) rhamnose state. Given that the variants of this class have reduced protein levels there may be a bigger effect activation in (-) rhamnose state relative to the (+) rhamnose state.

Another class of variants that had wild type activation levels both (+) and (-) rhamnose were also found. 20 out of the total 46 unique variants were identified to be in this class and these were spread across many residues in the Arm region. Interestingly, all the variants isolated from position L38 (5 out of 5) belongs to this class and suggests that this residue is not involved in making any inter-domain interactions.

Western blots to quantify the protein levels of variants. In order to understand whether changes in the activation levels (-) and (+) rhamnose had anything to do with the RhaR protein levels, I performed Western blots to quantify the levels of RhaR protein. I picked variants that represented all the different types of variants isolated during mutagenesis and tested for protein levels under both (-) and (+) rhamnose conditions. Protein levels are presented as a percentage relative to the amount of wild-type RhaR. Among the class of variants with wild type activation levels, two variants were tested for RhaR levels and found that these variants had no defects in protein levels both (-) and (+) rhamnose (Table 12). A total of five variants with increased activation (-) rhamnose and wild type activation (+) rhamnose were tested (L35D, L35K, L37R, D41S and F42I) and found that none of them had increased protein levels (-) rhamnose, suggesting that increased protein levels were not the cause of the increased activation (-) rhamnose (Table 12). This further supports our conclusion about the involvement of residue L35 in inhibitory inter-domain interactions. A total of four variants (K36R, K39T, D40G and F43R) were tested in the class of variants with decreased activation (-) rhamnose and wild type activation (+) rhamnose. Consistent with our previous hypothesis that these variants may have defects in protein stability, we found all variants had decreased protein levels (-) rhamnose, although one variant had reduced protein levels (+) rhamnose (Table 12).

Table 11. Transcription activation of *rhaS-lacZ* fusion by RhaR proposed Arm variants.

Mutation	% WT Activation[a]		Category[b]	Mutation	% WT Activation[a]		Category[b]
	(-) Rha	(+) Rha			(-) Rha	(+) Rha	
L35I	122±5	98±11	~WT	K39T	54±6	102±4	Down (-)
L35Y	131±10	94±4	~WT	K39S	94±6	82±16	~WT
L35D	143±12	93±9	Up (-)	K39V	110±25	128±17	~WT
L35G	143±7	94±5	Up (-)	K39A	162±16	122±2	Up (-)
L35V	171±13	96±3	Up (-)	D40G	13±22	79±1	Down (-)
L35K	180±8	99±6	Up (-)	D40E	85±9	134±15	~WT
K36A	52±23	87±14	Down (-)	D41V	108±3	124±3	~WT
K36R	64±15	88±23	Down (-)	D41A	168±9	148±28	Up
K36V	71±17	91±18	~WT	D41S	178±10	99±9	Up (-)
L37G	34±5	76±15	Down (-)	F42D	15±9	107±40	Down (-)
L37C	36±2	97±9	Down (-)	F42P	37±23	110±35	Down (-)
L37A	79±5	108±16	~WT	F42N	135±9	89±40	~WT
L37I	81±10	96±9	~WT	F42I	215±29	88±19	Up (-)
L37S	90±7	83±15	~WT	F43C	12±19	38±30	Down
L37M	125±14	98±11	~WT	F43R	13±6	85±4	Down (-)
L37R	192±14	75±20	Up (-)	F43A	15±7	96±13	Down (-)
L38G	63±17	101±7	~WT	F43G	15±18	101±5	Down (-)
L38R	69±8	10±6	~WT	F43V	148±1	75±4	Up (-)
L38W	83±7	105±4	~WT	A44G	18±3	43±6	Down
L38S	105±12	105±2	~WT	A44E	105±4	90±4	~WT
L38V	112±17	88±1	~WT	S45V	13±9	61±1	Down (-)

K39E	39±38	102±7	Down (-)	S45E	85±6	80±6	~WT
K39P	53±9	89±18	Down (-)	S45T	14±5	72±17	~WT

[a]β-Galactosidase activity was assayed from a single-copy Φ(*rhaS-lacZ*)Δ85 fusion in a strain with Δ(*rhaSR*)::*kan recA*::*cat* and wild type or RhaR variants expressed from a plasmid pHG165. Cultures were grown with or without L–rhamnose. Variants were assayed in groups with a wild-type RhaR activity range of 1.5-1.9 Miller Units (-) Rha and 8.5-9.0 Miller Units (+) Rha. Error is shown as the standard deviation converted to percent of the Miller Unit values.

[b]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.

Table 12. Protein levels of RhaR Arm variants.

Mutation	Catageory[a]	% WT[b] protein	
		(-) rhamnose	(+) rhamnose
L35K	Up (-)	79±19	68±3
L35D	Up (-)	112±18	114±3
K36R	Down (-)	36±3	83±1
L37R	Up (-)	37±4	86±2
L37M	~WT	138±1	139±1
L38S	~WT	116±6	94±2
K39T	Down (-)	31±8	25±0
D40G	Down (-)	14±2	2±0
D41S	Up (-)	31±1	40±3
F42I	Up (-)	85±1	95±0
F43R	Down (-)	23±3	19±1

[b]RhaR levels were estimated for different classes of variants by Western blotting and presented as a relative percentage of wild type.

[a]Category of variants: Down (-), reduced activity (-) Rha, near wild type (+) Rha; ~WT, near wild type activity (-) and (+) Rha; Up (-), increased activity (-) Rha, wild type activity (+) Rha.

Chapter 7

DISCUSSION

AraC family proteins are transcriptional regulators in many bacteria and are found to be essential for the expression of virulence factors required to cause diseases (24-28). With the ultimate goal of discovery of novel antibacterial agents that inhibit the AraC family proteins, here I have investigated the molecular mechanism of transcription activation by two AraC family activators Rns (virulence activators in Enterotoxigenic *Escherichia coli*) and RhaR (activator of L-rhamnose catabolic operons in *E. coli*). We expect that a better understanding of the molecular interactions involved in transcription activation could lead us to the rational design of the compounds that could target these AraC family proteins. In addition, we also tested a small molecule compound that was previously identified as an inhibitor of AraC family proteins RhaS and RhaR (activators of L-rhamnose catabolic operons in *E. coli*), for its ability to inhibit virulence regulators Rns and VirF.

Role of RS2 and AS2 regions in transcription activation by Rns. Rns is a key regulator in ETEC pathogenesis, because it activates the expression of major virulence factors called colonization factors (CFs) (41, 42). Although it was known that Rns activates expression of CFs required for successful ETEC colonization of the human gut, limited information has been gained on how Rns activates transcription of these genes. Mutational analysis by earlier researchers has found two mutations (I14T and N16D) in the Rns NTD that affected the transcription activity of Rns at its own promoter region, although these mutations have disparate effects on activation of other promoters (156). In addition, it was found that mutations in the HTH motifs of the Rns DBD inhibit Rns from binding to DNA and thereby affected transcription activation (93). Here, we have further expanded this analysis by assessing the role of a region of residues (RS2 region) in the NTD that are conserved among closely related Rns homologs and

their predicted contacting residues in the DBD (AS2 region) at both *cfaA* and modified *nlpA* promoter regions.

Our site-directed random mutagenesis has revealed the importance of RS2 residues in transcription activation. Despite the fact that the DBD of Rns is directly responsible for DNA binding and transcription activation, our analysis of variants for *cfaA-lacZ*$^{\Delta 60}$ expression and *nlpA*^{cfaA}*-lacZ* repression indicated that RS2 region residues N15, N16 and I17 in the NTD were required for the transcription activation function of Rns. Previous reports by Basturea *et al* showed that N16 residue is required for DNA binding function of Rns at *rns* promoter region (156). However, their analysis of N16D variant at *nlpA-lacZ* and *cfaA-lacZ* strains and my subsequent analysis of the same variant at *cfaA-lacZ*$^{\Delta 60}$ and *nlpA*^{cfaA}*-lacZ* suggested that N16D affects the transcription activation function by Rns. The impairment of transcription activation function by N16D supported our observation that the residue N16 is required for transcription activation function by Rns. Our site-directed mutagenesis of residues I12 and M18 resulted in inconclusive results, as protein levels were defective for most variants at these positions. Given the defects in protein levels for majority of variants at I12 and M18, these residues may be required for maintaining the stability of Rns. Residues K13 does not seem to be required for DNA binding and transcription activation function of Rns.

Our working model is that Rns residues N15, N16, I17 and possibly I14 in the RS2 region interact with the residues in the AS2 region and that the interactions between these two regions may hold the DBD in a conformation that enables it to activate transcription (most likely by direct interaction with RNAP). These kind of inter-domain interactions have previously been identified for the AraC family proteins AraC and ToxT (79, 101). For ToxT, interactions between the NTD and DBD have been identified in the presence of effector palmitoleic acid and

these interactions hold ToxT in a closed conformation such that ToxT does not activate transcription (79). The crystal structure of ToxT showed that in the effector-bound state, residues in each of the first two β-strands (β1, residues 7-14; β2, residues 19-26) are positioned to contact the ToxT-DBD residues (79). Most importantly, the residues N15, N16 and I17 of Rns align with [based on ClustalW analysis (Fig. 8) and I-TASSER predicted structures (157)] ToxT residues N23, N24 and L25 that are part of a ToxT β2 strand that contacts the ToxT DBD. Given the high sequence similarity of ToxT and Rns at these positions, I- TASSER structural prediction of N15, N16 and I17 of Rns in making inter domain interactions, and the involvement of N15, N16 and I17 in transcription activation function of Rns, we hypothesize that Rns N15, N16 and I17 residues may interact with the Rns AS2 region residues to activate transcription. In contrast to ToxT, no effector ligand has been identified for Rns. Therefore in the absence of effector, it is possible that the Rns NTD and DBD interact to form a constitutive open conformation to activate transcription and the mutations isolated in N15, N16 and I17 might have affected those interactions.

In order to test the possibility of the inter-domain interactions hypothesis, we mutagenized residues in the Rns AS2 region. Rns AS2 region residues are located in the middle of α7 helix and in a region from the end of α9 helix through the loop between α9 and α10 helix, and into α10 helix (α7, α9 – loop - α10). According to Rns I-TASSER structural model (157), these AS2 region residues are in an analogous position to residues in ToxT-DBD that are positioned to contact β2 sheet of ToxT NTD. Among the four residues mutagenized in the α7 helix of the Rns-DBD, variants at three residues, Q211, S215 and L219, suggested that these positions are not involved in either DNA binding or transcription activation function of Rns. Nonetheless, residue K216 was found to be required for transcription activation function of Rns.

In AraC, a number of AraC residues in α7 have been predicted to contact the AraC Arm in the absence of arabinose (101). In ToxT, the residue that aligns with Rns K216 (ToxT Y224), is adjacent to residue S223 (within α7 helix), that is positioned to make inter-domain interaction in the presence of effector (79). Although majority of the mutagenized residues in the α7 helix of Rns-DBD were not required for Rns activity, we found at least one residue K216 involved in the transcription activation function of Rns.

Variants isolated from Y251 were defective for *cfaA-lacZ*$^{\Delta 60}$ expression activation but not for repression, indicating that this residue is important for transcription activation function of Rns (Table 6). From residue G252, variants defective for *cfaA-lacZ*$^{\Delta 60}$ expression and variants defective for both *cfaA-lacZ*$^{\Delta 60}$ expression and *nlpA*^{cfaA}*-lacZ* repression were isolated, suggesting that G252 is required for both DNA binding and transcription activation function by Rns. I-TASSER prediction (157) of Rns structure showed that Rns residues Y251 and G252 are in the loop that follows the α9 helix that is located near the RS2 region residues. In the predicted structure, these two residues (Y251 and G252) are on the surface of the DBD that faces NTD, suggesting that they may play a role in imparting the structural conformation adopted by the DBD through interactions with NTD. Our results suggested that both Y251 and G252 are required for the transcription activation function of Rns. Residues Y251 and G252 are highly conserved in many AraC family proteins (Fig. 40) (35, 177, 237) and found to be important for transcription activity. For example, in PerA, the equivalent of Y251 residue, PerA Y255, is required for expression of fimbrial genes (92). In ToxT, the equivalent of Y251 (ToxT M259) and G252 (ToxT N260) are required for expression of cholerae toxin and toxin co-regulated pilus genes (238). In addition, in ToxT, the equivalent residue of Rns Y251 (ToxT M259) is positioned to enclose the effector binding pocket and predicted to contact residues in the NTD

and maintain the tertiary structure of ToxT (79, 238). The equivalent of Rns G252 in ToxT (ToxT N260), was a surface-exposed residue and predicted to be required for inter domain interactions within ToxT and interactions with RNA polymerase. Given the above evidence, we propose that Y251 and G252 are involved in inter-domain interactions with RS2 region residues, and these interactions are required for the transcription activation function of Rns (likely involving interaction with RNAP).

Mutagenesis of Rns residue H250 revealed a role for this residue in influencing both DNA binding and transcription activation. I-TASSER structural prediction of Rns showed that this residue is at the C-terminal end of the α9 helix (second helix of HTH2 motif or DNA recognition helix) and close to the RS2 region residues. Co-crystal structure of MarA with DNA showed that the residue corresponding to Rns H250 in MarA (MarA Y101) is at the C-terminal end of the DNA recognition helix of HTH2. (91). Although MarA Y101 was not found to be involved in making direct DNA interactions, positively charged residues before Y101 were involved in making direct contacts with DNA back bone. Given this, it is possible that H250 may contact the DNA directly. In addition, an equivalent residue of Rns H250 in ToxT (ToxT T258) was found to be involved in inter domain interactions (238). Therefore, we hypothesize that Rns H250 may participate in inter domain interactions and these interactions may keep the H250 in a conformation to either directly contact the DNA or make interactions required for transcription activation function of Rns.

In summary, our study expanded the knowledge with respect to the requirement of inter-domain interactions in Rns dependent expression from *cfaA* promoter region. We specifically identified residues in the Rns RS2 (N15, N16 and I17) and AS2 regions (K216, H250, Y251 and G252) that are required for Rns activity (Fig. 41). Our results indicate that Rns residues N15,

N16 and I17 are required for transcription activation function. Residues K216 and Y251 are required for transcription activation and residues H250 and G252 are required for both DNA binding and transcription activation function by Rns. The NTD residues (N15, N16 and I17) and the DBD residues (K216, H250, Y251 and G252) are located close to each other in the predicted Rns structure and a computational program ResMap (239) that predicts the inter-domain interactions predicted contacts between these residues in the structure model. ResMap predicted inter-domain interactions of N15 with H250; N16 with H250; and I17 with K216, H250 and Y251 (Table 13). In addition, we also found a very high sequence identity of these residues among the closely related homologs of Rns. Rns residues N15, N16, I17, H250, G252 showed 100% identity, while K216 and Y251 showed 100% similarity among the closely related homologs of Rns. This further suggested that these residues may have a conserved functional role in Rns activity. Among these seven residues that are predicted to be involved in making inter-domain interactions, a contact propensity matrix (240) showed high probability of interactions between side chains of Aspargine (N) with side chains of Histidine (H), Glycine (G) and Tyrosine (Y). This suggests a possibility of direct interactions between N15, N16 of RS2 region and H250, Y251 and G252 of AS2 region.

Although we could not conclude the role of I14 residue in Rns activity (because of isolation of a single variant I14W at this position), ResMap also predicted contacts between I14 and H250, Y251 and G252. In addition, previous reports by Basturea et al showed that I14T is required for making base specific DNA contacts by Rns at *rns* promoter region (156). However, I14T did not show any effect on Rns dependent expression from *cfaA* and *nlpA* promoter regions. Given the predicted interactions of I14 with AS2 residues, isolation of I14W that was defective for both DNA binding and transcription activation function, and previous observation of

involvement of I14T in DNA binding function of Rns at *rns* promoter region, we hypothesize that I14 residue may also be required for Rns both DNA binding and transcription activation function of Rns.

Overall, we propose that at least one mechanism of transcription activation in Rns involves inter-domain interactions between N15, N16, I17 and possibly I14 of NTD and K216, H250, Y251, G252 of DBD, and these interactions may impart the structure needed by Rns to activate transcription. In future, we would like to identify any specific interactions between these residues through screening for second-site-suppressors.

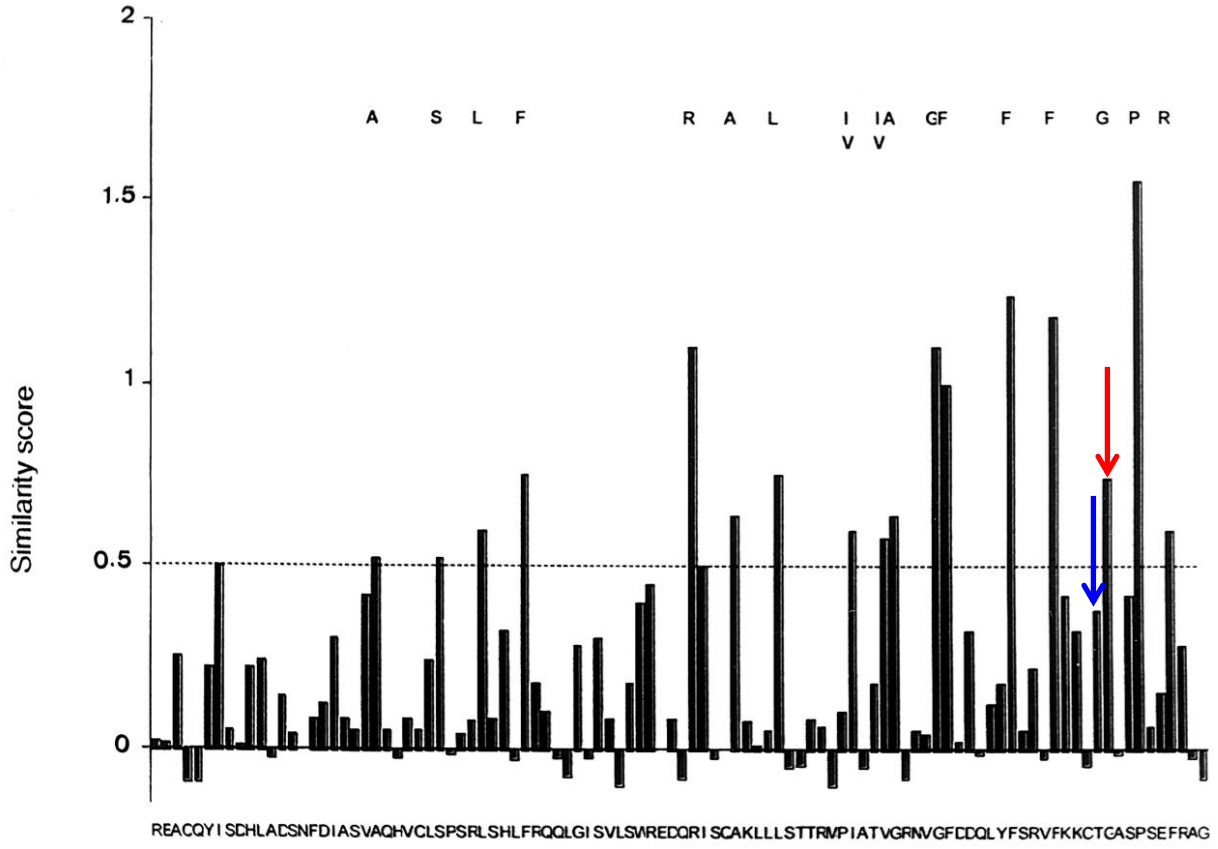


Fig. 40. Sequence similarity comparison of residues in the conserved DBD of AraC family proteins. Residues with a similarity score of above 0.5 (dotted line) represent the conserved residues among AraC family proteins. Residue corresponding to the Rns G252 had a similarity score of above 0.5 and was shown using red arrow. Residue corresponding to the Rns Y251 had a similarity score near to 0.5 and was shown using blue arrow. Modified from reference (36).

Table 13. ResMap prediction of inter domain interactions.

Rns residues found important	**Predicted contacts in other Rns domain**
I14	**H250**, **Y251***, **G252**
N15	K249*, **H250***
N16	E207*, Q211*, K249*, **H250***
I17	Q211, S215, **K216**, **H250**, **Y251***
K216	**N15***, **N16***, **I17**, H20
H250	**I14**, **N15***, **N16***, **I17**
Y251	I12*, K13*, **I14***, **N15***, **I17***
G252	K13, **I14**, **N15***, **N16***

Computational prediction program ResMap (239) was used to predict the inter domain interactions from the I-TASSER predicted structure of Rns (done by Dr. Susan Egan). The predicted contacted residues marked with an asterisk (*) have contact propensities equal to or greater than 1.0 for the residue they are predicted contact, suggesting a relatively high probability of an interaction (240).

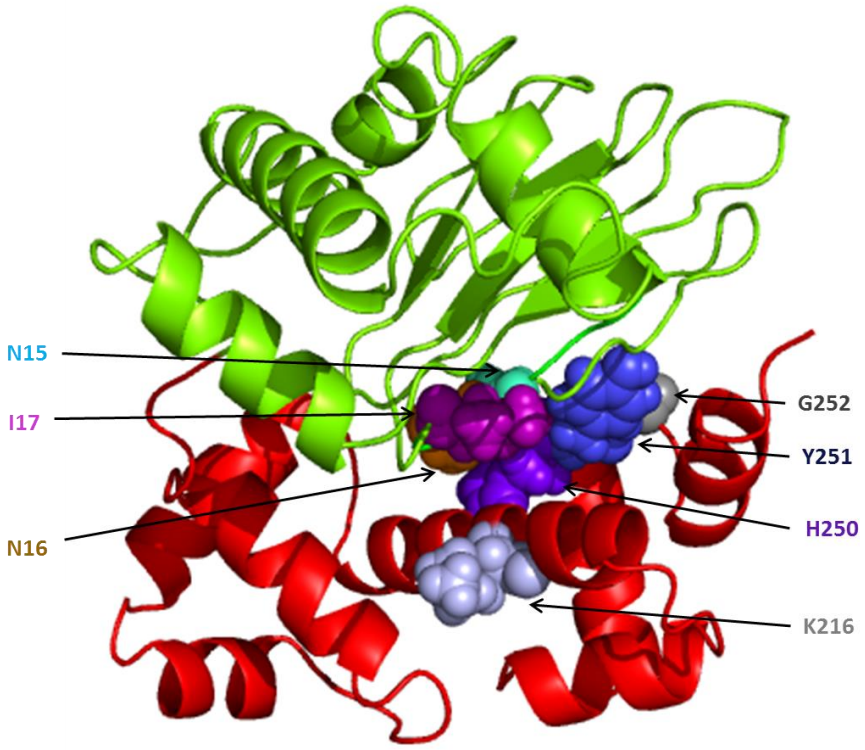


Fig. 41. RS2 and AS2 regions residues that are required for transcription activation by Rns. I-TASSER (157) was used to predict the 3D structure of Rns. NTD was shown in green color and DBD was shown in red color. Residues that were found to be essential for transcription in both RS2 and AS2 regions were shown as spheres and labeled in different colors as pointed with black arrows in the figure.

Purification and properties of Rns. Among AraC family proteins, only a fraction of the proteins have been characterized biochemically as the solubility of the proteins in this family tends to be very limited. To circumvent the problem of insolubility, I used GB1 and MBP solubility enhancement tags and purified Rns-DBD (residues 157-265) with GB1-His6 tag and full-length Rns with MBP tag. Overexpression of the Rns-DBD-GB1-His6 fusion at 15^{o}C has solubilized ~50% of the overexpressed protein. When the protein was purified using a Ni^{+} affinity chromatography column, we obtained a fairly high amount of soluble protein (~11mg for 100 ml initial culture). Later we performed EMSAs to test whether Rns-DBD can bind to DNA *in vitro.* Previously, DNaseI foot printing experiments by Basturea *et al* (156) using an MBP-Rns (128-265) construct that included DBD has shown that this construct did not bind to the *cfaA* promoter region even at high concentrations of 2.5 µM. Surprisingly, our EMSA experiments with the purified Rns-DBD-GB1-His6 showed that this protein can bind to *cfaA* promoter region even at concentrations of 1 µM. However, even at highest protein concentration tested (9 µM), only 7% of the DNA was bound by the protein. We expect that the differences in the sensitivities of the techniques may have led to observation of these varied results. DNase foot printing is more sensitive than EMSA in terms of using P^{32} labeled DNA as opposed to IR labeled DNA in our EMSAs. However, it is harder to notice a 7% DNA binding in DNase foot printing as we look for decrease in intensity of the DNA band by 7% in foot printing (100% to 93%) compared to gaining a new band in EMSA (0% to 7%).

Overexpression and purification of full-length Rns with MBP tag (attached at the N-terminus of Rns) showed that, as previously reported (156), the fusion protein is soluble, active and functional *in vitro*. Aggregation is a major problem of Rns and Rns aggregates readily when cleaved from MBP tag, and in elution buffer at concentrations above 2 mg/ml even without

cleavage. Aggregation was minimized by buffer exchange of protein into Tris buffer with reduced salt and no maltose (15 mM Tris-base, 75 mM KCl, 5 mM β-mercaptoethanol, 10% Glycerol, pH 7.6), and protein was successfully concentrated in this buffer to 20 mg/ml with minimal aggregation. Gel filtration analysis showed that the fusion protein elutes at the size of ~75 kDa indicating that Rns behaves as a monomer *in vitro*.

SE-1 as a small molecule inhibitor of AraC family activators. Many bacterial pathogens that pose severe health threats require AraC family transcriptional regulators to cause disease (24-28, 36). Given that AraC family proteins are required for the expression of virulence genes in multiple bacteria, they were thought to be excellent targets for the development of novel anti-bacterial agents (19-28, 36). AraC family virulence activators are non-essential for the growth of the bacteria and therefore targeting these proteins may not put pressure on the bacteria to develop resistance (9, 16-18). Dr. Susan Egan's lab recently identified an inhibitor, SE-1, which inhibited transcription activation by the AraC family proteins RhaS and RhaR (involved in activation of sugar rhamnose catabolic operons) by blocking their ability to bind to DNA (202). SE-1 inhibited RhaS and the RhaS DNA-binding domain (RhaS-DBD) to the same extent, suggesting that SE-1 acts on the structurally conserved DNA binding domain of the AraC family proteins (202). The goal of this study was to further test SE-1 potential to inhibit AraC family virulence activators VirF and Rns from *Shigella* and ETEC, respectively.

Using a cell-based reporter gene assay to test the ability of SE-1 to inhibit transcription activation by the VirF and Rns protein, we found that SE-1 decreased *virB-lacZ* reporter fusion expression in an *E. coli* strain with plasmid-expressed VirF (203) and Rns (done by Jeff Skredenske). The inhibition was dose-dependent and had an IC_{50} value of approximately 8 μM, suggesting that SE-1 inhibited transcription activation by VirF. An initial indication that this

inhibition was selective for Rns and VirF activity was the finding that SE-1 was substantially less inhibitory toward *lacZ* expression from a control fusion (*hts-lacZ*) than VirF and Rns dependent *lacZ* expression. The strains carrying the VirF-dependent and control *lacZ* fusions showed no detectable growth defects in the presence of SE-1, indicating that SE-1 did not have general toxicity toward these *E. coli* strains at the concentrations tested.

Given that SE-1 inhibition of the AraC family activators RhaS and RhaR involved blocking binding to DNA (202), we tested SE-1 for inhibition of DNA binding of purified VirF and Rns proteins to DNA carrying their respective binding sites, and found that SE-1 was able to fully inhibit DNA binding by Rns and VirF. We previously showed that SE-1 did not inhibit DNA binding by two proteins that are not related to the AraC family, LacI and CRP (202), suggesting that SE-1 is selective for AraC family proteins. Overall, our results support the hypothesis that the mechanism of action of SE-1 inhibition involves selectively blocking DNA binding by VirF and Rns by binding to the protein.

We next tested inhibition by SE-1 in *Shigella*, initially of endogenously expressed VirF in a *Shigella* strain carrying a *virB-lacZ* reporter construct. Similar to our findings in *E. coli*, we observed dose-dependent inhibition of VirF activation of *virB-lacZ* in *Shigella*. In addition, *Shigella* grown in the presence of SE-1 showed a decrease in transcription (assayed by qRT-PCR) of genes that are directly activated by VirF (*virB, icsA*) as well as genes that are directly activated by VirB (indirect VirF targets, *icsB, ipaB*). In these experiments, we observed up to two- to four-fold decreases in expression of these VirF-dependent genes, with values consistent with the *virB-lacZ* assays in *Shigella*. However, the same concentration of SE-1 resulted in lower levels of inhibition in *Shigella* than in *E. coli*, suggesting differences in uptake, efflux and/or stability of SE-1 in these two bacteria. Alternatively, similar to many commonly used antibiotics

(241) , there may be differences in the sensitivity to SE-1 in minimal media (used for *E. coli* growth) versus rich media (used for *Shigella* growth).

Several findings argue that the decreases in *virB-lacZ*, *virB, icsA, icsB*, and *ipaB* expression in *Shigella* were due to inhibition of VirF activity and not a global decrease in gene expression in the presence of SE-1. First, there was no detectable decrease in the growth rate of *Shigella* in the presence of up to 40 μM SE-1, whereas global decreases in gene expression of a comparable magnitude would be expected to affect the growth rate. Second, the qRT-PCR results were normalized to two different internal control genes (which would be expected to exhibit decreased expression if there was a global decrease in gene expression). Normalization of each of the internal control genes relative to the other showed that there was no significant change in the expression of the control genes in the presence of 40 μM SE-1. Finally, the magnitude of the inhibition of *virB* expression by SE-1 detected in the *virB-lacZ* assays and the qRT-PCR assays were essentially the same, indicating that the normalization of the qRT-PCR results to the internal controls did not alter the extent of the inhibition (as would have been expected if expression of the internal control genes was also inhibited).

Finally, we investigated the effect of SE-1 on the ability of *Shigella* to invade mouse fibroblast (L-929) cells in tissue culture. A dose-dependent decrease in invasion was observed with a maximum inhibition of 70% at 40 μM inhibitor. This inhibition required pre-incubation of *Shigella* with SE-1, consistent with the hypothesis that SE-1 did not destabilize preassembled T3SS or other *Shigella* virulence components, inhibit post-transcriptional processes in *Shigella*, or cause cytotoxicity to the L-929 cells at the concentrations tested. The latter is further supported by the finding that SE-1 did not result in a detectable decrease in metabolism of L-929 cells in AlamarBlue®assays at these same concentrations. Overall, our results suggest a model in

which SE-1 inhibits VirF-dependent transcription activation of *Shigella* virulence genes, including at least *icsA*, *virB*, *icsB* and *ipaB*. The decrease in invasion we detected is consistent with and can be attributed to the effect of decreased expression of these virulence genes in the VirF regulon. The two- to four-fold decreases in expression of these genes appear to be sufficient to decrease invasion of *Shigella* by 70%, compared to the uninhibited invasion level. The *ipaB* gene exhibited the largest decrease in expression, four-fold, suggesting that reduction in IpaB expression may have limited invasion in the presence of SE-1. This is a reasonable hypothesis given the central role IpaB plays in *Shigella* invasion. IpaB is a component of the translocon pores in the host cell membrane (179) that enable translocation of *Shigella* effectors into host cells (182), ultimately modulating host cell cytoskeletal dynamics and leading to bacterial invasion (180). In addition, only a subset of the VirF-regulated virulence genes that are required for invasion were assayed (177, 178), thus, decreased expression of other genes likely also contributed to the invasion defects. Overall, SE-1 is non-cytotoxic toward the eukaryotic cells tested (at concentrations up to 40 μM) and non-bactericidal toward *E. coli* and *Shigella* (at concentrations up to 40 μM) and has potential to be developed into a novel anti-bacterial agent. However, SE-1 does show evidence of non-specific inhibition at higher concentrations (partial inhibition of *hts-lacZ* at higher concentrations and inhibition of *Shigella* growth at 80 μM SE-1), and its current potency is not sufficient for therapeutic applications. Thus, chemical optimization of SE-1 will be needed to improve both its potency and its specificity before it can be considered a lead compound for further development.

We have now shown that SE-1 selectively inhibits DNA binding by four different AraC family proteins RhaS, RhaR (202), Rns and VirF. The VirF protein shares only about 15% amino acid sequence identity and 40% similarity with RhaS and RhaR (the pairwise RhaS-VirF and

RhaR-VirF comparisons are nearly the same) (Fig. 42). Rns shares only very limited sequence identity with RhaS (8%) and RhaR (12%), and relatively high identity with VirF (35%). Given our prior finding that SE-1 blocks transcription activation by the RhaS DBD to at least the same extent as the full-length RhaS protein (202), we hypothesize that the DNA binding domain is the likely site of action of SE-1. The sequence comparisons in the conserved DNA binding domain are somewhat higher than the full-length proteins, but are still relatively low, with the VirF DBD sharing about 22% identity and 52% similarity with the RhaS and RhaR DBDs (the pairwise comparison of Rns DBD with the RhaS and RhaR DBDs are nearly the same). Despite having these sequences differences, the DNA binding of all these proteins were inhibited by SE-1. This finding supports the hypothesis that SE-1 may bind to a relatively conserved region of AraC family DNA binding domains. In addition, our computational docking with BSP-SLIM (227) and Swiss-Dock (www.swissdock.ch, Molecular Modeling Group, Swiss Institute of Bioinformatics, Lausanne, Switzerland) predicted that SE-1 can bind between the two HTH motifs of the predicted DBD structures of RhaS, RhaR, Rns and VirF, and thus could sterically inhibit the binding of these proteins to their respective promoter DNA.

We predict that SE-1 can inhibit other AraC family virulence activators in addition to RhaS, RhaR, VirF and Rns. Recently, our lab also showed that SE-1 can inhibit DNA binding by AraC family transcription activator ToxT from *Vibrio cholerae*. The ability of SE-1 to inhibit several AraC family activators led us to hypothesize that SE-1 can bind to a region that shares high structural and likely sequence similarity among AraC family proteins. Our docking results of SE-1 with the DBD of RhaS predicted 10 residues in the DBD (V188, W190, D191, L220, N221, R224, F248, S249, H253 and L257) (performed by Jeff Skredenske and Dr. Susan Egan) that could potentially interact with SE-1. In addition, docking of SE-1 on to the predicted

structures of Rns DBD and VirF DBD also showed interactions of SE-1 at positions that align with RhaS DBD. Sequence comparison of RhaS, RhaR, Rns, VirF and ToxT showed sequence similarities (ClustalW) in a range of 60-80% (60% for ToxT and 80% for Rns and VirF) at these 10 positions relative to the most common residue found at each position (Table 14). Among these 10 positions, 8 of the positions share sequence similarity among this set of proteins, and align with RhaS V188, L190, L220, R224, F248, S249, H253 and L257. The high sequence similarities in RhaS, RhaR, Rns, VirF and ToxT proteins among these 8 residues suggests that the mechanism of inhibition of these proteins by SE-1 may involve contacts with these residues.

In order to find whether SE-1 has potential to inhibit many other AraC family proteins, I searched for the sequence similarities in the predicted SE-1 interaction residues among 197 AraC family proteins that were previously aligned by Dr. Susan Egan. Our results indicated a substantially higher sequence similarity in five out of the total eight predicted SE-1 interaction residues. These include residues that align with RhaS V188, L190, L220, R224 and F248. The residue at the position of RhaS V188 shared 72% similarity (24% identity), L190 shared 93% similarity (46% identity), L220 shared 90% similarity (62% identity), R224 shared 99% similarity (88% identity) and F248 shared 100% similarity (69% identity) among these 197 AraC family proteins (Table 15). In addition, the residue that aligns with RhaS H253 also shared 43% sequence identity, but had less similarity relatively to other positions. The high level of sequence similarity among the residues predicted to make interactions with SE-1 suggests that the compound can inhibit many other AraC family proteins in addition to RhaS, RhaR, Rns, VirF and ToxT, and could be developed into a broad spectrum antibiotic. We also compared sequences of 197 proteins to find the proteins that have least sequence similarities in the predicted SE-1 binding sites. We expect that identification of these proteins would give us an

idea of the percentage of proteins that are not inhibited by SE-1. Our results only showed one protein, YeaM, which had completely different sequences at three of the five positions that are highly conserved (RhaS V188, R224 and F248) among the predicted eight SE-1 interaction residues (Table 16).

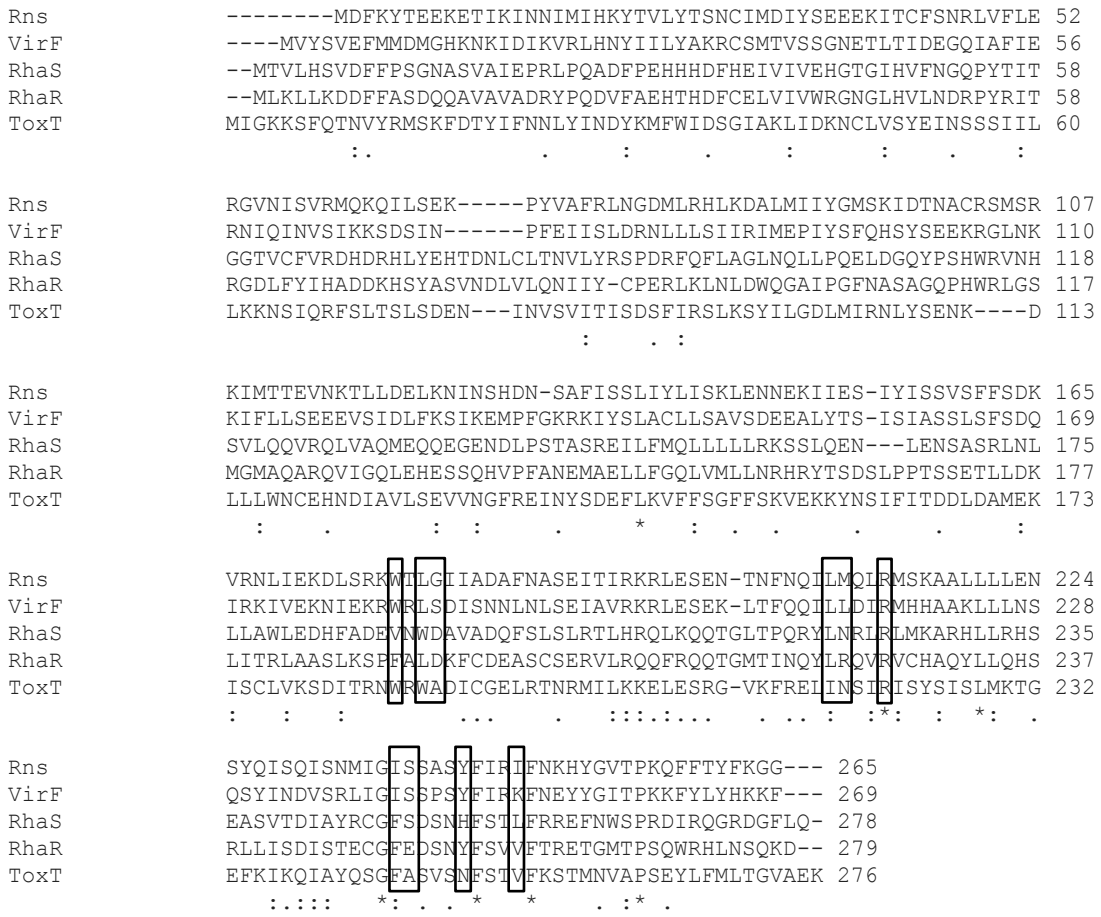


Fig. 42. Alignment of Rns, VirF, ToxT, RhaS and RhaR . Sequences were aligned using ClustalW method (157). Residues that are predicted to be contacted by SE-1 are shown in black boxes. Gaps in the alignment are marked "-".

Table 14. Molecular docking prediction of SE-1 interaction of RhaS, RhaR, Rns, VirF and ToxT.

Protein	188[a]	190[a]	191[a]	220[a]	221[a]	224[a]	248[a]	249[a]	253[a]	257[a]	I[b]	S[c]	D[d]
RhaS	V	W	D	L	N	R	F	S	H	L	4	3	3
RhaR	F	L	D	L	R	R	F	E	Y	V	5	2	3
Rns	W	L	G	L	M	R	I	S	Y	I	6	2	2
VirF	W	L	S	L	L	R	I	S	Y	K	6	1	3
ToxT	W	W	A	I	N	R	F	A	N	V	3	3	4

[a]Position of SE-1 interaction residues in RhaS

[b] Number of residues identical to the most common residue

[c]Number of residues similar to the most common residue

[d]Number of residues different from the most common residue

The most common residue at each position among the five proteins was labeled red.

Table 15. Sequence similarity comparison among the predicted SE-1 interaction residues in 197 AraC family proteins.

Residue	188[a]	190[a]	220[a]	224[a]	248[a]	249[a]	253[a]	257[a]
V	47	27	9					61
L	40	91	122				2	
W	42	33						
F	34	2	1		136			
I	19	23	45		8			7
M		7	1		1			1
E								
Q						2		
H				3				
K				20				
Y					52		84	
T						2		
N						7		
A			2				6	25
M						68	11	
S								
R				173				
Percent Similarity	**72**	**93**	**90**	**99**	**100**	**40**	**52**	**54**
Percent Identity	**24**	**46**	**62**	**88**	**69**	**35**	**43**	**31**

[a]Position of the predicted SE-1 interaction residues in RhaS

Total number of proteins with residues identical to the most common residue at each position are labeled red.

Total number of proteins with residues similar to the most common residue at each position are labeled green.

Table 16. Sequence alignment of the predicted SE-1 interaction residues in 197 AraC family proteins.

Protein	188	190	220	224	248	249	253	257
AARP_PROST/22-120	L	L	I	R	F	D	S	R
ADA_ECOLI/85-183	V	L	Q	R	F	P	S	K
ADA_MYCTU/87-185	D	V	A	R	F	S	Q	T
ADAA_BACSU/102-200	L	L	I	R	I	A	Y	L
ADIY_ECOLI/149-246	W	L	I	R	Y	N	Y	V
AGGR_ECOLX/164-261	W	L	L	R	F	S	Y	L
APPY_ECOLI/139-236	W	L	L	R	Y	N	Y	A
ARAC_CITFR/180-279	F	I	R	R	F	D	Y	V
ARAC_ECO57/180-279	F	I	R	R	F	D	Y	V
ARAC_ECOLI/180-279	F	I	R	R	F	D	Y	V
ARAC_ERWCH/186-284	L	I	R	R	Y	D	Y	V
ARAC_SALTY/180-279	F	I	R	R	F	D	Y	V
ARACL_STRAT/202-300	W	V	L	R	Y	G	A	A
ARACL_STRLI/202-300	W	V	L	R	Y	G	A	A
CAF1R_YERPE/8-107	I	I	I	R	Y	D	T	E
CFAD_ECOLX/164-261	W	L	L	R	I	S	Y	V
CHBR_ECOLI/168-274	S	L	I	R	Y	S	L	T
CSVR_ECOLX/166-263	W	L	L	R	I	S	Y	I
ENVY_ECOLI/149-246	W	L	V	R	Y	S	Y	V
EUTR_ECOLI/243-344	V	V	L	R	F	W	Q	D
EUTR_SALTY/243-344	L	V	L	R	F	W	Q	D
EXSA_PSEAE/171-269	W	L	I	R	F	S	Y	S
FAPR_ECOLX/154-251	W	L	I	R	Y	T	Y	T
FEAR_ECOLI/199-299	L	P	I	R	F	S	H	V
GADW_ECO57/139-236	W	L	L	R	Y	S	Y	T
GADW_ECOL6/139-236	W	L	L	R	Y	S	Y	T
GADW_ECOLI/139-236	W	L	L	R	Y	S	Y	T
GADW_SHIFL/139-236	W	L	L	R	Y	S	Y	T
GADX_ECO27/145-242	W	L	L	R	Y	H	Y	V
GADX_ECO57/145-242	W	L	L	R	Y	H	Y	V
GADX_ECOL6/145-242	W	L	L	R	Y	H	Y	V
GADX_ECOLI/145-242	W	L	L	R	Y	H	Y	V
GADX_SHIFL/145-242	W	L	L	R	Y	H	Y	V
GLXA_RHIME/223-321	L	R	Y	R	F	S	H	A
HILD_SALTY/209-306	W	L	Y	R	Y	D	Y	C
HRPB_RALSO/375-477	L	T	I	R	I	R	A	G
INVF_SALTI/112-210	N	M	L	R	Y	S	H	E
INVF_SALTY/112-210	N	M	L	R	Y	S	H	E
LACR_STAXY/174-272	I	V	L	R	Y	K	L	N
LCRF_YERPE/167-265	W	L	I	R	F	S	Y	S
LUMQ_PHOLE/148-246	I	V	V	R	F	S	S	A
MARA_ECO57/12-110	L	L	I	K	F	E	T	T

MARA_ECOLI/12-110	L	L	I	K	F	E	T	T
MARA_SALEN/12-110	L	L	I	K	F	E	T	T
MARA_SALTI/12-110	L	L	I	K	F	E	T	T
MARA_SALTY/12-110	L	L	I	K	F	E	T	T
MARA_SHIFL/12-110	L	L	I	K	F	E	T	T
MELR_ECOL6/194-292	L	I	I	R	F	R	R	T
MELR_ECOLI/194-292	L	I	I	R	F	R	R	T
MMSR_PSEAE/201-299	L	L	F	K	Y	D	Y	L
MSMR_STRMU/176-274	L	V	I	R	F	S	A	A
MXIE_SHIFL/99-199	V	I	L	R	Y	A	H	E
MXIE_SHISO/99-199	V	I	L	R	Y	A	H	E
ORUR_PSEAE/241-338	P	L	L	R	F	N	N	A
PCHR_PSEAE/201-296	P	L	L	R	Y	S	H	A
PERA_ECO27/168-265	W	L	M	K	F	N	Y	V
POCR_SALTY/195-293	L	L	V	R	F	S	Y	V
PQRA_PROVU/7-107	I	I	V	R	F	S	T	I
RAFR_PEDPE/176-274	C	I	L	R	Y	K	T	A
RAMA_ENTCL/9-107	L	I	I	K	F	E	T	I
RAMA_KLEPN/9-107	L	I	I	K	F	D	T	V
RHAR_ECO24/179-277	F	L	L	R	F	E	Y	V
RHAR_ECO57/179-277	F	L	L	R	F	E	Y	V
RHAR_ECOHS/179-277	F	L	L	R	F	E	Y	V
RHAR_ECOK1/179-277	F	L	L	R	F	E	Y	V
RHAR_ECOL5/179-277	F	L	L	R	F	E	Y	V
RHAR_ECOL6/179-277	F	L	L	R	F	E	Y	V
RHAR_ECOLI/179-277	F	L	L	R	F	E	Y	V
RHAR_ECOUT/179-277	F	L	L	R	F	E	Y	V
RHAR_ENT38/179-277	F	L	L	R	F	E	Y	V
RHAR_ENTS8/179-277	F	L	L	R	F	E	Y	V
RHAR_ERWCT/181-279	F	F	L	R	F	E	Y	V
RHAR_KLEP7/178-276	F	L	L	R	F	E	Y	V
RHAR_MANSM/174-272	F	L	L	R	Y	S	Y	V
RHAR_SALAR/179-277	F	L	L	R	F	E	Y	V
RHAR_SALCH/179-277	F	L	L	R	F	E	Y	V
RHAR_SALPA/179-277	F	L	L	R	F	E	Y	V
RHAR_SALPB/179-277	F	L	L	R	F	E	Y	V
RHAR_SALTI/179-277	F	L	L	R	F	E	Y	V
RHAR_SALTY/179-277	F	L	L	R	F	E	Y	V
RHAR_SHIBS/179-277	F	L	L	R	F	E	Y	V
RHAR_SHIDS/179-277	F	L	L	R	F	E	Y	V
RHAR_SHIF8/179-277	F	L	L	R	F	E	Y	V
RHAR_SHIFL/179-277	F	L	L	R	F	E	Y	V
RHAR_YERP3/179-277	F	M	L	R	F	E	Y	V
RHAR_YERPA/179-277	F	M	L	R	F	E	Y	V
RHAR_YERPE/179-277	F	M	L	R	F	E	Y	V
RHAR_YERPN/179-277	F	M	L	R	F	E	Y	V
RHAR_YERPP/179-277	F	M	L	R	F	E	Y	V
RHAR_YERPS/179-277	F	M	L	R	F	E	Y	V

RHAS_CITK8/174-272	V	W	L	R	F	G	H	L
RHAS_ECO24/174-272	V	W	L	R	F	S	H	L
RHAS_ECO57/174-272	V	W	L	R	F	S	H	L
RHAS_ECOHS/174-272	V	W	L	R	F	S	H	L
RHAS_ECOK1/174-272	V	W	L	R	F	S	H	L
RHAS_ECOL5/174-272	V	W	L	R	F	S	H	L
RHAS_ECOL6/174-272	V	W	L	R	F	S	H	L
RHAS_ECOLC/174-272	V	W	L	R	F	S	H	L
RHAS_ECOLI/174-272	V	W	L	R	F	S	H	L
RHAS_ECOUT/174-272	V	W	L	R	F	S	H	L
RHAS_ENT38/174-272	V	W	L	R	F	G	H	L
RHAS_ERWCT/175-273	V	W	L	R	F	S	H	Q
RHAS_KLEP7/174-272	I	W	L	R	F	G	H	L
RHAS_MANSM/171-268	I	W	V	R	F	N	Y	C
RHAS_SALAR/174-272	V	W	L	R	F	G	H	L
RHAS_SALCH/174-272	V	W	L	R	F	G	H	L
RHAS_SALPA/174-272	V	W	L	R	F	G	H	L
RHAS_SALPB/174-272	V	W	L	R	F	G	H	L
RHAS_SALTI/174-272	V	W	L	R	F	G	H	L
RHAS_SALTY/174-272	V	W	L	R	F	G	H	L
RHAS_SHIBS/174-272	V	W	L	R	F	S	H	L
RHAS_SHIDS/174-272	V	W	L	R	F	S	H	L
RHAS_SHIF8/174-272	V	W	L	R	F	S	H	L
RHAS_SHIFL/174-272	V	W	L	R	F	S	H	L
RHAS_SHISS/174-272	V	W	L	R	F	S	H	L
RHAS_YERP3/174-272	V	W	L	R	F	G	H	L
RHAS_YERPA/174-272	V	W	L	H	F	G	H	L
RHAS_YERPE/174-272	V	W	L	H	F	G	H	L
RHAS_YERPN/174-272	V	W	L	H	F	G	H	L
RHAS_YERPP/174-272	V	W	L	R	F	G	H	L
RHAS_YERPS/174-272	V	W	L	R	F	G	H	L
RHRA_RHIME/210-310	F	I	V	R	F	N	Y	T
RIPA_CORDI/119-216	T	L	R	R	F	S	S	A
RIPA_COREF/112-209	T	L	R	R	F	S	S	A
RIPA_CORGL/112-209	T	L	R	R	F	A	S	A
RNS_ECOLX/164-261	W	L	L	R	I	S	Y	I
ROB_ECO57/8-106	L	L	I	R	F	D	T	A
ROB_ECOLI/8-106	L	L	I	R	F	D	T	A
ROB_SHIFL/8-106	L	L	I	R	F	D	T	A
SIRC_SALTI/195-292	W	Q	Y	R	Y	D	Y	I
SIRC_SALTY/195-292	W	Q	Y	R	Y	D	Y	I
SOXS_ECO57/8-106	L	I	I	R	Y	V	T	V
SOXS_ECOL6/8-106	L	I	I	R	Y	V	T	V
SOXS_ECOLI/8-106	L	I	I	R	Y	V	T	V
SOXS_SALTY/8-106	L	I	I	R	Y	V	T	V
TCPN_VIBC3/172-269	W	W	I	R	F	A	Y	V
TCPN_VIBCH/172-269	W	W	I	R	F	A	Y	V
TETD_ECOLX/31-129	L	L	I	R	F	D	S	R

THCR_RHOER/227-328	L	V	L	R	F	L	R	E
URER_ECOLX/171-268	W	L	L	R	F	G	Y	V
URER_PROMI/171-268	W	L	L	R	F	G	Y	A
VIRF_SHIDY/161-258	W	L	L	R	I	S	Y	K
VIRF_SHIFL/161-258	W	L	L	R	I	S	Y	K
VIRF_SHISO/161-258	W	L	L	R	I	S	Y	K
VIRF_YERE8/167-265	W	L	I	R	F	S	Y	S
VIRF_YEREN/167-265	W	L	I	R	F	S	Y	S
VIRS_MYCTU/236-334	C	A	I	R	Y	S	A	S
VQSM_PSEAE/226-323	P	L	V	R	Y	S	S	A
XYLR_ECO57/288-386	I	V	I	K	Y	P	Y	V
XYLR_ECOL6/288-386	I	V	I	K	Y	P	Y	V
XYLR_ECOLI/288-386	I	V	I	K	Y	P	Y	V
XYLR_HAEIN/288-386	I	V	I	K	Y	P	Y	V
XYLS_PSEPU/214-315	I	L	I	K	F	L	R	N
XYLS1_PSEPU/214-315	I	L	I	K	F	L	R	N
XYLS2_PSEPU/39-140	I	L	I	K	F	L	R	K
XYLS3_PSEPU/214-315	I	L	I	K	F	F	R	N
XYLS4_PSEPU/214-315	I	L	I	K	F	F	R	N
Y077_STAAW/158-256	L	L	L	R	F	S	S	T
Y078_STAAS/158-256	L	L	L	R	F	S	S	T
Y084_STAAC/158-256	L	L	L	R	F	S	S	T
Y097_STAAN/158-256	L	L	L	R	F	S	S	T
Y101_STAAM/158-256	L	L	L	R	F	S	S	T
Y1052_HAEIN/194-295	W	I	L	R	Y	Q	H	V
Y107_STAAR/158-256	L	L	L	R	F	S	S	T
Y1395_MYCTU/242-343	P	F	L	R	Y	T	T	A
Y1430_MYCBO/242-343	P	F	L	R	Y	T	T	A
Y161_STAAB/153-250	V	V	L	R	F	S	M	H
Y165_STAES/152-249	V	V	L	R	F	S	M	H
Y184_STAA8/153-250	V	V	L	R	F	S	M	H
Y198_STAAS/153-250	V	V	L	R	F	S	M	H
Y198_STAAW/153-250	V	V	L	R	F	S	M	H
Y201_STAAC/153-250	V	V	L	R	F	S	M	H
Y214_STAAR/153-250	V	V	L	R	F	S	M	H
Y215_STAAN/153-250	V	V	L	R	F	S	M	H
Y217_STAA3/153-250	V	V	L	R	F	S	M	H
Y223_STAAM/153-250	V	V	L	R	F	S	M	H
Y2406_STAEQ/152-249	V	V	L	R	F	S	M	H
Y4FK_RHISN/318-417	V	I	L	R	F	S	R	L
YBBB_BACSU/166-264	I	L	I	R	Y	Q	Y	I
YBCM_ECOLI/165-262	W	A	L	R	Y	K	R	R
YCGK_ALTCA/67-163	I	I	I	R	F	N	Y	V
YDEO_ECO57/137-233	W	L	L	R	Y	A	Y	A
YDEO_ECOL6/137-233	W	L	L	R	Y	A	Y	A
YDEO_ECOLI/137-233	W	L	L	R	Y	A	Y	A
YDEO_SHIFL/137-233	W	L	L	R	Y	A	Y	A
YDIP_ECOLI/183-281	I	L	V	R	Y	E	H	L

YEAM_ECOLI/158-258	G	L	R	Q	Y	D	A	M
YFIF_BACSU/192-289	L	L	V	K	F	S	Y	V
YHTH1_STAAU/158-256	L	L	L	R	F	S	S	T
YIDL_ECOLI/197-295	W	V	Y	R	F	F	H	A
YIJO_ECOLI/172-270	L	R	L	R	F	V	Y	L
YISR_BACSU/183-281	T	I	T	R	M	E	Y	L
YKGA_ECOLI/19-117	I	I	I	R	F	D	S	E
YKGD_ECOLI/177-278	W	V	L	R	Y	A	S	A
YMCR_STRLA/184-281	R	V	A	R	F	A	H	D
YPDC_ECOLI/184-282	L	R	V	R	F	P	Y	V
YQHC_ECOLI/213-311	L	V	L	R	Y	E	Q	E

Residues similar to the most common residue at each position are shown in colored boxes.

(Original alignment of these proteins was provided by Dr. Susan Egan).

Role of RhaR Arm in L-rhamnose response. Given that RhaR DBD activates transcription in response to binding of L-rhamnose at the N-terminal domain, we sought to identify how the L-rhamnose signal is transmitted from NTD to the DBD. We earlier discovered that deletion of RhaR extension (first 34 amino acids) does not play any role in transmitting this signal. However, we could not identify the role of RhaR residues that align with the AraC Arm because of defects in protein levels in the RhaR variants with deletions in the aligned region (200).

In the present study, I mutagenized residues in the RhaR Arm region (residues 35-45) by site-directed random mutagenesis and isolated a total of 46 unique variants. Unlike the instability of protein truncations, we expected that single mutations may be less likely to cause defects in protein stability, and thus enable functional analysis of the Arm region. Our results showed that none of the variants had decreased activity in the presence of rhamnose, suggesting that this region may not be involved in making stimulatory inter-domain contacts (+) rhamnose. We found several variants that had increased activity (-) rhamnose. Most of the variants in position L35 had this phenotype [increased activation (-) rhamnose without increased protein levels] suggesting that this residue may be involved in making inhibitory inter-domain contacts in the (-) rhamnose state of wild-type RhaR. In AraC, the equivalent residue of RhaR L35 (AraC D7) is the first residue in the AraC Arm that makes inhibitory inter-domain interactions with AraC-DBD in the (-) arabinose state (100). Our results identified one residue, L35, in the RhaR Arm that is involved in keeping the protein in a non-activation state in the absence of rhamnose through making inhibitory inter-domain interactions.

Our mutagenesis also resulted in identification of variants with increased activation in (-) rhamnose state for residues L37, K39, F42 and F43. However, variants with this phenotype were unique at their respective residues, suggesting that those variants are probably gain-of-residue-

function variants (They may not be a loss-of-residue-function variants because for a loss-of-residue-function we expect the majority of variants to have increased activation (-) rhamnose at these positions). Although these residues may not make direct inhibitory interactions with the DBD in (-) rhamnose state, they may be close enough to the DBD, such that some variants could make inhibitory contacts with the DBD or alter the conformation of other residues in the NTD to contact the DBD.

Overall, our site-directed random mutagenesis of the RhaR Arm region suggested that residue L35 makes inhibitory contacts until the availability of L-rhamnose to the cells. We hypothesize that RhaR Arm might make inter-domain interaction with the residues in the DBD in (-) rhamnose state and those interactions might inhibit RhaR activation. In order to find the residues in the DBD that may contact the Arm residues, another graduate student Bria Kettle mutagenized residues in a region (allosteric site 2, AS2) that were predicted to be near the RhaR Arm region. Mutagenesis of residues in the AS2 region resulted in identification of variants at H269 with increased activation in (-) rhamnose state, suggesting the involvement of H269 in inhibitory inter-domain contacts in (-) rhamnose state. Further, she found the potential interactions between L35 and H269 by screening for second-site-suppressors that restore the RhaR activation levels to wild type in the absence of rhamnose. The structural model of RhaR (Fig. 43) also predicted inter-domain interactions between H269 and L35. These results suggested to us that interactions between residues L35 and H269 in the absence of rhamnose may keep the protein in a non-activating state. Previously, Ana Kolin from our lab found several variants in a RhaR residue T279 with increased activation in (-) rhamnose state (200). This residue T279 is on the surface of the RhaR DNA binding domain and in a region (between D276 and D285), which was shown to be involved in making contacts with σ70 of RNAP (81, 200).

These results led us to hypothesize that the residue T279 may interfere with RNAP contacts in the (-) rhamnose state and thus keep the RhaR activity to basal levels. Given the current and previous reports, we propose that in the absence of rhamnose, inhibitory inter-domain interactions between L35 and H269 may alter the conformation of RhaR DBD such that T279 interferes with the RNAP contacts made by RhaR. However, availability of rhamnose may lead to loss of inhibitory inter-domain interactions between L35 and H269, and thus allow the residues in the region around H269 (probably D276) to make interactions with RNAP.

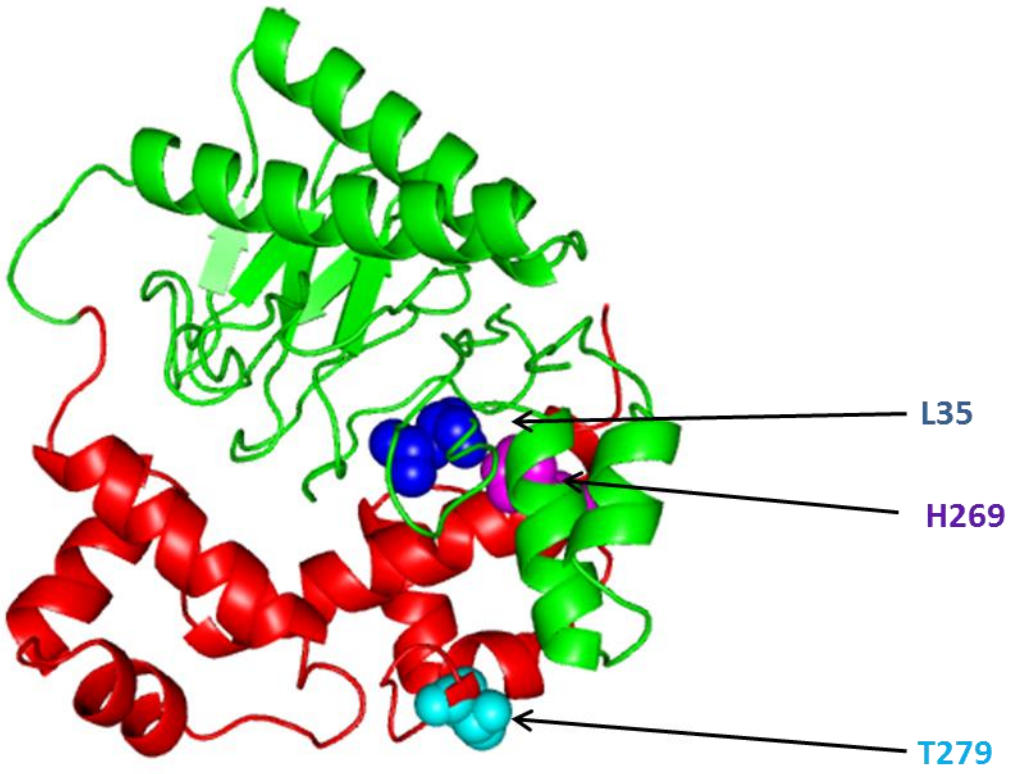


Fig. 43. I-TASSER prediction of RhaR structure. I-TASSER (157) was used to predict the 3D structure of RhaR. NTD was shown in green and DBD was shown in red. Residue T279 was shown as sphere in cyan. Residues L35 and H269 that were found to make inhibitory inter domain interactions were shown as blue and purple spheres, respectively.

REFERENCES

1. **Fleming A.** 2001. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. 1929. Bull. World Health Organ. **79:**780-790.
2. **Coates A, Hu Y, Bax R, Page C.** 2002. The future challenges facing the development of new antimicrobial drugs. Nat. Rev. Drug Discov. **1:**895-910.
3. **Pickering LK.** 2004. Antimicrobial resistance among enteric pathogens. Semin. Pediatr. Infect. Dis. **15:**71-77.
4. **Walsh C.** 2003. Where will new antibiotics come from? Nat. Rev. Microbiol. **1:**65-70.
5. **Projan SJ, Bradford PA.** 2007. Late stage antibacterial drugs in the clinical pipeline. Curr. Opin. Microbiol. **10:**441-446.
6. **Talbot GH, Bradley J, Edwards JE, Jr., Gilbert D, Scheld M, Bartlett JG, Antimicrobial Availability Task Force of the Infectious Diseases Society of A.** 2006. Bad bugs need drugs: an update on the development pipeline from the Antimicrobial Availability Task Force of the Infectious Diseases Society of America. Clin. Infect. Dis. **42:**657-668.
7. **Spellberg B, Guidos R, Gilbert D, Bradley J, Boucher HW, Scheld WM, Bartlett JG, Edwards J, Jr., Infectious Diseases Society of A.** 2008. The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. Clin. Infect. Dis. **46:**155-164.
8. **Walsh C.** 2003. Antibiotics : actions, origins, resistance. ASM Press, Washington, D.C.
9. **Rasko DA, Sperandio V.** 2010. Anti-virulence strategies to combat bacteria-mediated disease. Nature reviews. Drug discovery **9:**117-128.
10. **Alksne LE.** 2002. Virulence as a target for antimicrobial chemotherapy. Expert Opin Investig Drugs **11:**1149-1159.
11. **Alksne LE, Burgio P, Hu W, Feld B, Singh MP, Tuckman M, Petersen PJ, Labthavikul P, McGlynn M, Barbieri L, McDonald L, Bradford P, Dushin RG, Rothstein D, Projan SJ.** 2000. Identification and analysis of bacterial protein secretion inhibitors utilizing a *secA-lacZ* reporter fusion system. Antimicrob. Agents Chemother. **44:**1418-1427.

12. **Chan PF, Macarron R, Payne DJ, Zalacain M, Holmes DJ.** 2002. Novel antibacterials: a genomics approach to drug discovery. Curr. Drug Targets Infect. Disord. **2:**291-308.
13. **Donadio S, Carrano L, Brandi L, Serina S, Soffientini A, Raimondi E, Montanini N, Sosio M, Gualerzi CO.** 2002. Targets and assays for discovering novel antibacterial agents. J. Biotechnol. **99:**175-185.
14. **Fritz B, Raczniak GA.** 2002. Bacterial genomics: potential for antimicrobial drug discovery. Biodrugs **16:**331-337.
15. **Isaacson RE.** 2002. Genomics and the prospects for the discovery of new targets for antibacterial and antifungal agents. Curr. Pharm. Des. **8:**1091-1098.
16. **Hughes D.** 2003. Exploiting genomics, genetics and chemistry to combat antibiotic resistance. Nat. Rev. Genet. **4:**432-441.
17. **Knowles DJ.** 1997. New strategies for antibacterial drug design. Trends Microbiol. **5:**379-383.
18. **Schmidt FR.** 2004. The challenge of multidrug resistance: actual strategies in the development of novel antibacterials. Appl. Microbiol. Biotechnol. **63:**335-343.
19. **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.
20. **Hung DT, Shakhnovich EA, Pierson E, Mekalanos JJ.** 2005. Small-molecule inhibitor of *Vibrio cholerae* virulence and intestinal colonization. Science **310:**670-674.
21. **Shakhnovich EA, Hung DT, Pierson E, Lee K, Mekalanos JJ.** 2007. Virstatin inhibits dimerization of the transcriptional activator ToxT. Proc. Natl. Acad. Sci. U. S. A. **104:**2372-2377.
22. **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.
23. **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.

24. **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.
25. **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.
26. **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.
27. **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.
28. **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.
29. **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.
30. **Sakai T, Sasakawa C, Makino S, Yoshikawa M.** 1986. DNA sequence and product analysis of the virF locus responsible for congo red binding and cell invasion in *Shigella flexneri* 2a. Infect. Immun. **54:**395-402.
31. **Tobe T, Schoolnik GK, Sohel I, Bustamante VH, Puente JL.** 1996. Cloning and characterization of *bfpTVW*, genes required for the transcriptional activation of *bfpA* in enteropathogenic *Escherichia coli*. Mol. Microbiol. **21:**963-975.
32. **Nataro JP, Yikang D, Yingkang D, Walker K.** 1994. AggR, a transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative *Escherichia coli*. J. Bacteriol. **176:**4691-4699.

33. **de Haan LA, Willshaw GA, van der Zeijst BA, Gaastra W.** 1991. The nucleotide sequence of a regulatory gene present on a plasmid in an Enterotoxigenic *Escherichia coli* strain of serotype O167:H5. FEMS Microbiol. Lett. **67:**341-346.

34. **Klaasen P, de Graaf FK.** 1990. Characterization of FapR, a positive regulator of expression of the 987P operon in Enterotoxigenic *Escherichia coli* Mol. Microbiol. **4:**1779-1783.

35. **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.

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

37. **Perez-Rueda E, Collado-Vides J.** 2000. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. Nucleic Acids Res. **28:**1838-1847.

38. **Parker LL, Hall BG.** 1990. Characterization and nucleotide sequence of the cryptic cel operon of *Escherichia coli* K12. Genetics **124:**455-471.

39. **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.

40. **Caron J, Scott JR.** 1990. A *rns*-like regulatory gene for colonization factor antigen I (CFA/I) that controls expression of CFA/I pilin. Infect. Immun. **58:**874-878.

41. **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.* . Proc. Natl. Acad. Sci. U. S. A. **86:**963-967.

42. **Bodero MD, Harden EA, Munson GP.** 2008. Transcriptional regulation of subclass 5b fimbriae. BMC Microbiol. **8:**180.

43. **Fetherston JD, Bearden SW, Perry RD.** 1996. YbtA, an AraC-type regulator of the Yersinia pestis pesticin/yersiniabactin receptor. Mol. Microbiol. **22:**315-325.

44. **Schleif R.** 2010. AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. FEMS Microbiol. Rev. **34:**779-796.

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

46. **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.

47. **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.

48. **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.

49. **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.

50. **Gomez-Duarte OG, Kaper JB.** 1995. A plasmid-encoded regulatory region activates chromosomal *eaeA* expression in enteropathogenic *Escherichia coli*. Infect. Immun. **63:**1767-1776.

51. **Hart E, Yang J, Tauschek M, Kelly M, Wakefield MJ, Frankel G, Hartland EL, Robins-Browne RM.** 2008. RegA, an AraC-like protein, is a global transcriptional regulator that controls virulence gene expression in *Citrobacter rodentium*. Infect. Immun. **76:**5247-5256.

52. **Kane CD, Schuch R, Day WA, Jr., Maurelli AT.** 2002. MxiE regulates intracellular expression of factors secreted by the *Shigella flexneri* 2a type III secretion system. J. Bacteriol. **184:**4409-4419.

53. **Hulbert RR, Taylor RK.** 2002. Mechanism of ToxT-dependent transcriptional activation at the *Vibrio cholerae tcpA* promoter. J. Bacteriol. **184:**5533-5544.

54. **Hovey AK, Frank DW.** 1995. Analyses of the DNA-binding and transcriptional activation properties of ExsA, the transcriptional activator of the *Pseudomonas aeruginosa* exoenzyme S regulon. J. Bacteriol. **177:**4427-4436.

55. **D'Orazio SEF, Collins CM.** 1993. The plasmid-encoded urease gene cluster of the family *Enterobacteriaceae* is positively regulated by UreR, a member of the AraC family of transcriptional activators. J. Bacteriol. **175:**3459-3467.

56. **Kaniga K, Bossio JC, Galan JE.** 1994. The *Salmonella typhimurium* invasion genes *invF* and *invG* encode homologues of the AraC and PulD family of proteins. Mol. Microbiol. **13:**555-568.

57. **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.

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

59. **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.

60. **Gendlina I, Gutman DM, Thomas V, Collins CM.** 2002. Urea-dependent signal transduction by the virulence regulator UreR. J. Biol. Chem. **277:**37349-37358.

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

62. **Englesberg E.** 1965. Isolation of Mutants in the L-arabinose Gene-enzyme Complex. In Methods in Enzymology, ed. **9:**15-21.

63. **Steele MI, Lorenz D, Hatter K, Sokatch JR.** 1992. Characterization of the *mmsAB* operon of *Pseudomonas aeruginosa* PAO encoding methylmalonate-semialdehyde dehydrogenase and 3-hydroxyisobutarate dehydrogenase. J. Biol. Chem. **267:**13585-13592.

64. **Spooner R. A. LK, 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.

65. **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.

66. **Skarstad K, Thony B, Hwang DW, Kornberg A.** 1993. A novel binding protein of the origin of the *Escherichia coli* chromosome. J. Biol. Chem. **268:**5365-5370.

67. **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.

68. **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.

69. **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. Proc. Natl. Acad. Sci. U. S. A. **82:**2688-2692.

70. **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.

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

72. **Eustance RJ, Schleif RF.** 1996. The linker region of AraC protein. J. Bacteriol. **178:**7025-7030.

73. **Seedorff J, Schleif R.** 2011. Active role of the interdomain linker of AraC. J. Bacteriol. **193:**5737-5746.

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

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

76. **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.

77. **Bustos SA, Schleif RF.** 1993. Functional domains of the AraC protein. Proc. Natl. Acad. Sci. USA **90:**5638-5642.

78. **Prouty MG, Osorio CR, Klose KE.** 2005. Characterization of functional domains of the *Vibrio cholerae* virulence regulator ToxT. Mol. Microbiol. **58:**1143-1156.

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. Proc. Natl. Acad. Sci. U. S. A. **107:**2860-2865.

80. **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.

81. **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.

82. **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.

83. **Lonetto MA, Rhodius V, Lamberg K, Kiley P, Busby S, Gross C.** 1998. Identification of a contact site for different transcription activators in region 4 of the *Escherichia coli* RNA polymerase s^{70} subunit. J. Mol. Biol. **284:**1353-1365.

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

85. **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.

86. **Zhang X, Schleif R.** 1998. Catabolite gene activator protein mutations affecting activity of the araBAD promoter. J. Bacteriol. **180:**195-200.

87. **Brutinel ED, Vakulskas CA, Yahr TL.** 2009. Functional domains of ExsA, the transcriptional activator of the *Pseudomonas aeruginosa* type III secretion system. J. Bacteriol. **191:**3811-3821.

88. **Kaldalu N, Toots U, de Lorenzo V, Ustav M.** 2000. Functional domains of the TOL plasmid transcription factor XylS. J. Bacteriol. **182:**1118-1126.

89. **Gambino L, Gracheck SJ, Miller PF.** 1993. Overexpression of the MarA positive regulator is sufficient to confer multiple antibiotic resistance in *Escherichia coli*. J. Bacteriol. **175:**2888-2894.

90. **Wu J, Weiss B.** 1991. Two divergently transcribed genes, *soxR* and *soxS*, control a superoxide response regulon of *Escherichia coli*. J. Bacteriol. **173:**2864-2871.

91. **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. Proc. Natl. Acad. Sci. USA **95:**10413-10418.

92. **Ibarra JA, Garcia-Zacarias CM, Lara-Ochoa C, Carabarin-Lima A, Tecpanecatl-Xihuitl JS, Perez-Rueda E, Martinez-Laguna Y, Puente JL.** 2013. Further characterization of functional domains of PerA, role of amino and carboxy terminal domains in DNA binding. PLoS One **8:**e56977.
93. **Mahon V, Smyth CJ, Smith SG.** 2010. Mutagenesis of the Rns regulator of Enterotoxigenic *Escherichia coli* reveals roles for a linker sequence and two helix-turn-helix motifs. Microbiology **156:**2796-2806.
94. **Bhende PM, Egan SM.** 1999. Amino acid-DNA contacts by RhaS: an AraC family transcription activator. J. Bacteriol **181:**5185-5192.
95. **Niland P, Huhne R, Muller-Hill B.** 1996. How AraC interacts specifically with its target DNAs. J. Mol. Biol. **264:**667-674.
96. **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.
97. **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.
98. **Steffen D, Schleif.** 1977. Overproducing AraC Protein with Lambda-arabinose Transducing Phage. Mol. Gen. Genet. **157:**333-339.
99. **Englesberg E, Irr J, Power J, Lee N.** 1965. Positive Control of Enzyme Synthesis by Gene C in the L-arabinose System. J. Bacteriol. **90:**946-957.
100. **Saviola B, Seabold R, Schleif RF.** 1998. Arm-domain interactions in AraC. J. Mol. Biol. **278:**539-548.
101. **Wu M, Schleif R.** 2001. Mapping arm-DNA-binding domain interactions in AraC. J. Mol. Biol. **307:**1001-1009.
102. **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.
103. **Saviola B, Seabold R, Schleif RF.** 1998. Arm-domain interactions in AraC. J Mol Biol **278:**539-548.

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

105. **Harmer T, Wu M, Schleif R.** 2001. The role of rigidity in DNA looping-unlooping by AraC. Proc. Natl. Acad. Sci. U. S. A. **98:**427-431.

106. **Wu M, Schleif R.** 2001. Strengthened arm-dimerization domain interactions in AraC. The Journal of biological chemistry **276:**2562-2564.

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

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

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

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

111. **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. Proc. Natl. Acad. Sci. U. S. A. **81:**5017-5020.

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

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

114. **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.

115. **Diemert DJ.** 2006. Prevention and self-treatment of traveler's diarrhea. Clin. Microbiol. Rev. **19:**583-594.

116. **Black RE.** 1993. Persistent diarrhea in children of developing countries. Pediatr. Infect. Dis. J. **12:**751-761; discussion 762-754.

117. **WHO.** 2006. Future directions for research on Enterotoxigenic *Escherichia coli* vaccines for developing countries. Wkly. Epidemiol. Rec. **81:**97-104.

118. **de la Cabada Bauche J, Dupont HL.** 2011. New Developments in Traveler's Diarrhea. Gastroenterology & hepatology **7:**88-95.

119. **Beatty ME, Adcock PM, Smith SW, Quinlan K, Kamimoto LA, Rowe SY, Scott K, Conover C, Varchmin T, Bopp CA, Greene KD, Bibb B, Slutsker L, Mintz ED.** 2006. Epidemic diarrhea due to Enterotoxigenic *Escherichia coli*. . Clin. Infect. Dis. **42:**329-334.

120. **Daniels NA.** 2006. Enterotoxigenic *Escherichia coli* : traveler's diarrhea comes home. Clin. Infect. Dis. **42:**335-336.

121. **Beatty ME, Bopp CA, Wells JG, Greene KD, Puhr ND, Mintz ED.** 2004. Enterotoxin-producing *Escherichia coli* O169:H41, United States. Emerg. Infect. Dis. **10:**518-521.

122. **Black RE.** 1990. Epidemiology of travelers' diarrhea and relative importance of various pathogens. Rev. Infect. Dis. **12 Suppl 1:**S73-79.

123. **Turner SM, Chaudhuri RR, Jiang ZD, DuPont H, Gyles C, Penn CW, Pallen MJ, Henderson IR.** 2006. Phylogenetic comparisons reveal multiple acquisitions of the toxin genes by Enterotoxigenic *Escherichia coli* strains of different evolutionary lineages. J. Clin. Microbiol. **44:**4528-4536.

124. **Allen KP, Randolph MM, Fleckenstein JM.** 2006. Importance of heat-labile enterotoxin in colonization of the adult mouse small intestine by human Enterotoxigenic *Escherichia coli* strains. Infect. Immun. **74:**869-875.

125. **Dorsey FC, Fischer JF, Fleckenstein JM.** 2006. Directed delivery of heat-labile enterotoxin by Enterotoxigenic *Escherichia coli*. . Cell. Microbiol. **8:**1516-1527.

126. **Tsai SC, Noda M, Adamik R, Moss J, Vaughan M.** 1987. Enhancement of choleragen ADP-ribosyltransferase activities by guanyl nucleotides and a 19-kDa membrane protein. Proc. Natl. Acad. Sci. U. S. A. **84:**5139-5142.

127. **Sears CL, Kaper JB.** 1996. Enteric bacterial toxins: mechanisms of action and linkage to intestinal secretion. Microbiol. Rev. **60:**167-215.

128. **Hughes JM, Murad F, Chang B, Guerrant RL.** 1978. Role of cyclic GMP in the action of heat-stable enterotoxin of *Escherichia coli*. Nature **271:**755-756.

129. **Rao MC, Guandalini S, Smith PL, Field M.** 1980. Mode of action of heat-stable *Escherichia coli* enterotoxin. Tissue and subcellular specificities and role of cyclic GMP. Biochim. Biophys. Acta **632:**35-46.

130. **Chao AC, de Sauvage FJ, Dong YJ, Wagner JA, Goeddel DV, Gardner P.** 1994. Activation of intestinal CFTR Cl- channel by heat-stable enterotoxin and guanylin via cAMP-dependent protein kinase. EMBO J. **13:**1065-1072.

131. **Zafriri D, Oron Y, Eisenstein BI, Ofek I.** 1987. Growth advantage and enhanced toxicity of *Escherichia coli* adherent to tissue culture cells due to restricted diffusion of products secreted by the cells. J. Clin. Invest. **79:**1210-1216.

132. **Satterwhite TK, Evans DG, DuPont HL, Evans DJ, Jr.** 1978. Role of *Escherichia coli* colonisation factor antigen in acute diarrhoea. Lancet **2:**181-184.

133. **Evans DG, Silver RP, Evans DJ, Jr., Chase DG, Gorbach SL.** 1975. Plasmid-controlled colonization factor associated with virulence in *Esherichia coli* enterotoxigenic for humans. Infect. Immun. **12:**656-667.

134. **Gaastra W, Svennerholm AM.** 1996. Colonization factors of human Enterotoxigenic *Escherichia coli* (ETEC). Trends Microbiol. **4:**444-452.

135. **Peruski LF, Jr., Kay BA, El-Yazeed RA, El-Etr SH, Cravioto A, Wierzba TF, Rao M, El-Ghorab N, Shaheen H, Khalil SB, Kamal K, Wasfy MO, Svennerholm AM, Clemens JD, Savarino SJ.** 1999. Phenotypic diversity of Enterotoxigenic *Escherichia coli* strains from a community-based study of pediatric diarrhea in periurban Egypt. J. Clin. Microbiol. **37:**2974-2978.

136. **Grewal HM, Valvatne H, Bhan MK, van Dijk L, Gaastra W, Sommerfelt H.** 1997. A new putative fimbrial colonization factor, CS19, of human enterotoxigenic *Escherichia coli*. Infect. Immun. **65:**507-513.

137. **Pichel M, Binsztein N, Viboud G.** 2000. CS22, a novel human enterotoxigenic *Escherichia coli* adhesin, is related to CS15. Infect. Immun. **68:**3280-3285.

138. **Qadri F, Svennerholm AM, Faruque AS, Sack RB.** 2005. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. Clin. Microbiol. Rev. **18:**465-483.

139. **Evans DG, Evans DJ, Jr.** 1978. New surface-associated heat-labile colonization factor antigen (CFA/II) produced by Enterotoxigenic *Escherichia coli* of serogroups O6 and O8. Infect. Immun. **21:**638-647.

140. **Levine MM, Ristaino P, Marley G, Smyth C, Knutton S, Boedeker E, Black R, Young C, Clements ML, Cheney C, et al.** 1984. Coli surface antigens 1 and 3 of

colonization factor antigen II-positive Enterotoxigenic *Escherichia coli* : morphology, purification, and immune responses in humans. Infect. Immun. **44:**409-420.

141. **Smyth CJ.** 1982. Two mannose-resistant haemagglutinins on Enterotoxigenic *Escherichia coli* of serotype O6:K15:H16 or H-isolated from travellers' and infantile diarrhoea. J. Gen. Microbiol. **128:**2081-2096.
142. **McConnell MM, Thomas LV, Willshaw GA, Smith HR, Rowe B.** 1988. Genetic control and properties of coli surface antigens of colonization factor antigen IV (PCF8775) of Enterotoxigenic *Escherichia coli* Infect. Immun. **56:**1974-1980.
143. **Wolf MK, Andrews GP, Tall BD, McConnell MM, Levine MM, Boedeker EC.** 1989. Characterization of CS4 and CS6 antigenic components of PCF8775, a putative colonization factor complex from Enterotoxigenic *Escherichia coli* E8775. Infect. Immun. **57:**164-173.
144. **Froehlich B, Husmann L, Caron J, Scott JR.** 1994. Regulation of *rns*, a positive regulatory factor for pili of Enterotoxigenic *Escherichia coli* J. Bacteriol. **176:**5385-5392.
145. **Pilonieta MC, Bodero MD, Munson GP.** 2007. CfaD-dependent expression of a novel extracytoplasmic protein from Enterotoxigenic *Escherichia coli* J. Bacteriol. **189:**5060-5067.
146. **Munson GP, Holcomb LG, Alexander HL, Scott JR.** 2002. In vitro identification of Rns-regulated genes. J. Bacteriol. **184:**1196-1199.
147. **Munson GP, Scott JR.** 1999. Binding site recognition by Rns, a virulence regulator in the AraC family. J. Bacteriol. **181:**2110-2117.
148. **Tobe T, Nagai S, Okada N, Adler B, Yoshikawa M, Sasakawa C.** 1991. Temperature-regulated expression of invasion genes in *Shigella flexneri i*s controlled through the transcriptional activation of the *virB* gene on the large plasmid. Mol. Microbiol. **5:**887-893.
149. **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.
150. **Jordi BJ, Dagberg B, de Haan LA, Hamers AM, van der Zeijst BA, Gaastra W, Uhlin BE.** 1992. The positive regulator CfaD overcomes the repression mediated by

histone-like protein H-NS (H1) in the CFA/I fimbrial operon of *Escherichia coli*. EMBO J. **11:**2627-2632.

151. **Morin N, Tirling C, Ivison SM, Kaur AP, Nataro JP, Steiner TS.** 2010. Autoactivation of the AggR regulator of enteroaggregative Escherichia coli in vitro and in vivo. FEMS Immunol. Med. Microbiol. **58:**344-355.
152. **Stonehouse EA, Hulbert RR, Nye MB, Skorupski K, Taylor RK.** 2011. H-NS binding and repression of the ctx promoter in Vibrio cholerae. J. Bacteriol. **193:**979-988.
153. **Puente JL, Bieber D, Ramer SW, Murray W, Schoolnik GK.** 1996. The bundle-forming pili of enteropathogenic Escherichia coli: transcriptional regulation by environmental signals. Mol. Microbiol. **20:**87-100.
154. **Munson GP, Scott JR.** 2000. Rns, a virulence regulator within the AraC family, requires binding sites upstream and downstream of its own promoter to function as an activator. Mol. Microbiol. **36:**1391-1402.
155. **Mahon V, Fagan RP, Smith SG.** 2012. Snap denaturation reveals dimerization by AraC-like protein Rns. Biochimie **94:**2058-2061.
156. **Basturea GN, Bodero MD, Moreno ME, Munson GP.** 2008. Residues near the amino terminus of Rns are essential for positive autoregulation and DNA binding. J. Bacteriol. **190:**2279-2285.
157. **Zhang Y.** 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics **9:**40.
158. **Kotloff KL, Winickoff JP, Ivanoff B, Clemens JD, Swerdlow DL, Sansonetti PJ, Adak GK, Levine MM.** 1999. Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. Bull World Health Organ. **77:**651-666.
159. **Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omore R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acacio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM.** 2013. Burden and aetiology of

diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet. **382:**209-222.

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

161. **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.

162. **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.

163. **Porter ME, Dorman CJ.** 1997. Differential regulation of the plasmid-encoded genes in the *Shigella flexneri* virulence regulon. Mol. Gen. Genet. **256:**93-103.

164. **Porter ME, Dorman CJ.** 1994. A role for H-NS in the thermo-osmotic regulation of virulence gene expression in *Shigella flexneri*. J. Bacteriol. **176:**4187-4191.

165. **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.

166. **Durand JM, Dagberg B, Uhlin BE, Björk GR.** 2000. Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of *Shigella flexneri*: the expression of the *virF* gene. Mol. Microbiol. **35:**924-935.

167. **Porter ME, Dorman CJ.** 1997. Positive regulation of *Shigella flexneri* virulence genes by integration host factor. J. Bacteriol. **179:**6537-6550.

168. **Tobe T, Yoshikawa M, Sasakawa C.** 1995. Thermoregulation of *virB* transcription in *Shigella flexneri* by sensing of changes in local DNA superhelicity. J. Bacteriol. **177:**1094-1097.

169. **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.

170. **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.

171. **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.

172. **Goldberg MB, Theriot JA, Sansonetti PJ.** 1994. Regulation of surface presentation of IcsA, a *Shigella* protein essential to intracellular movement and spread, is growth phase dependent. Infect. Immun. **62:**5664-5668.

173. **Suzuki T, Miki H, Takenawa T, Sasakawa C.** 1998. Neural Wiskott-Aldrich syndrome protein is implicated in the actin-based motility of *Shigella flexneri*. EMBO J. **17:**2767-2776.

174. **Goldberg MB.** 1997. *Shigella* actin-based motility in the absence of vinculin. Cell Motil Cytoskeleton. **37:**44-53.

175. **Egile C, Loisel TP, Laurent V, Li R, Pantaloni D, Sansonetti PJ, Carlier MF.** 1999. Activation of the CDC42 effector N-WASP by the *Shigella flexneri* IcsA protein promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility. J. Cell Biol. **146:**1319-1332.

176. **Suzuki T, Mimuro H, Suetsugu S, Miki H, Takenawa T, Sasakawa C.** 2002. Neural Wiskott-Aldrich syndrome protein (N-WASP) is the specific ligand for *Shigella* VirG among the WASP family and determines the host cell type allowing actin-based spreading. Cell. Microbiol. **4:**223-233.

177. **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.

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

179. **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.

180. **Tran Van Nhieu G, Caron E, Hall A, Sansonetti PJ.** 1999. IpaC induces actin polymerization and filopodia formation during *Shigella* entry into epithelial cells. EMBO J. **18:**3249-3262.

181. **Chen Y, Smith MR, Thirumalai K, Zychlinsky A.** 1996. A bacterial invasin induces macrophage apoptosis by binding directly to ICE. EMBO J. **15:**3853-3860.

182. **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.

183. **Ogawa M, Yoshimori T, Suzuki T, Sagara H, Mizushima N, Sasakawa C.** 2005. Escape of intracellular *Shigella* from autophagy. Science. **307:**727-731.

184. **Kayath CA, Hussey S, El hajjami N, Nagra K, Philpott D, Allaoui A.** 2010. Escape of intracellular *Shigella* from autophagy requires binding to cholesterol through the type III effector, IcsB. Microbes Infect **12:**956-966.

185. **Sasakawa C, Kamata K, Sakai T, Makino S, Yamada M, Okada N, Yoshikawa M.** 1988. Virulence-associated genetic regions comprising 31 kilobases of the 230-kilobase plasmid in *Shigella flexneri* 2a. J. Bacteriol. **170:**2480-2484.

186. **Mills JA, Venkatesan MM, Baron LS, Buysse JM.** 1992. Spontaneous insertion of an IS1-like element into the *virF* gene is responsible for avirulence in opaque colonial variants of *Shigella flexneri* 2a. Infect. Immun. **60:**175-182.

187. **Caron J, Maneval DR, Kaper JB, Scott JR.** 1990. Association of *rns* homologs with colonization factor antigens in clinical *Escherichia coli* isolates. Infect. Immun. **58:**3442-3444.

188. **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.

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

190. **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.

191. **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.

192. **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.

193. **Tate CG, Muiry JAR, Henderson PJF.** 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. **287:**6923-6932.

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

195. **Wilson DM, Ajl S.** 1957. Metabolism of L-rhamnose by *Escherichia coli*. I. L-rhamnose isomerase. J. Bacteriol. **73:**410-414.

196. **Sawada H, Takagi Y.** 1964. The Metabolism of L-Rhamnose in *Escherichia coli*. 3. L-Rhamulose-Phosphate Aldolase. Biochim. Biophys. Acta **92:**26-32.

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

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

199. **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.

200. **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.

201. **Donadio S, Brandi L, Serina S, Sosio M, Stinchi S.** 2005. Discovering novel antibacterial agents by high throughput screening. Frontiers in Drug Design & Discovery **1:**3-16.

202. **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.

203. **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.

204. **Haldimann A, Wanner BL.** 2001. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J. Bacteriol. **183:**6384-6393.

205. **Stewart GSAB, 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.

206. **Casadaban M.** 1976. Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage Lambda and Mu. J. Mol. Biol. **104:**541-555.

207. **Silber KR, Keiler KC, Sauer RT.** 1992. Tsp: a tail-specific protease that selectively degrades proteins with nonpolar C termini. Proc. Natl. Acad. Sci. U. S. A. **89:**295-299.

208. **Miller JH.** 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

209. **Schneider CA, Rasband WS, Eliceiri KW.** 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods **9:**671-675.

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

211. **Runyen-Janecky L, Dazenski E, Hawkins S, Warner L.** 2006. Role and regulation of the *Shigella flexneri sit* and *mntH* systems. Infect. Immun. **74:**4666-4672.

212. **Broach WH, Egan N, Wing HJ, Payne SM, Murphy ER.** 2012. VirF-independent regulation of *Shigella virB* transcription is mediated by the small RNA RyhB. PLoS One. **7:**e38592.

213. **Bore E, Hebraud M, Chafsey I, Chambon C, Skjaeret C, Moen B, Moretro T, Langsrud O, Rudi K, Langsrud S.** 2007. Adapted tolerance to benzalkonium chloride in *Escherichia coli* K-12 studied by transcriptome and proteome analyses. Microbiology. **153:**935-946.

214. **Kosek M, Bern C, Guerrant RL.** 2003. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. Bull. World Health Organ. **81:**197-204.

215. **DuPont HL, Hornick RB, Snyder MJ, Libonati JP, Formal SB, Gangarosa EJ.** 1972. Immunity in shigellosis. I. Response of man to attenuated strains of *Shigella*. J. Infect. Dis. **125:**5-11.

216. **Mel DM, Terzin AL, Vuksic L.** 1965. Studies on vaccination against bacillary dysentery. 3. Effective oral immunization against *Shigella flexneri* 2a in a field trial. Bull. World Health Organ. **32:**647-655.

217. **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.

218. **Dentchev V, Marinova S, Vassilev T, Bratoyeva M, Linde K.** 1990. Live *Shigella flexneri* 2a and *Shigella sonnei* I vaccine candidate strains with two attenuating markers. II. Preliminary results of vaccination of adult volunteers and children aged 2-17 years. Vaccine. **8:**30-34.

219. **Alikhani MY, Hashemi SH, Aslani MM, Farajnia S.** 2013. Prevalence and antibiotic resistance patterns of diarrheagenic Escherichia coli isolated from adolescents and adults in Hamedan, Western Iran. Iranian journal of microbiology **5:**42-47.

220. **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.

221. **Rees DC, Congreve M, Murray CW, Carr R.** 2004. Fragment-based lead discovery. Nat. Rev. Drug Discov. **3:**660-672.

222. **Nguyen CC, Saier MH, Jr.** 1995. Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcription factors. FEBS Lett. **377:**98-102.

223. **Swint-Kruse L, Matthews KS.** 2009. Allostery in the LacI/GalR family: variations on a theme. Curr. Opin. Microbiol. **12:**129-137.

224. **Korner H, Sofia HJ, Zumft WG.** 2003. Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. FEMS Microbiol. Rev. **27:**559-592.

225. **Sakellaris H, Scott JR.** 1998. New tools in an old trade: CS1 pilus morphogenesis. Mol Microbiol. **30:**681-687.

226. **Anantha RP, McVeigh AL, Lee LH, Agnew MK, Cassels FJ, Scott DA, Whittam TS, Savarino SJ.** 2004. Evolutionary and functional relationships of colonization factor antigen I and other class 5 adhesive fimbriae of enterotoxigenic *Escherichia coli*. Infect Immun. **72:**7190-7201.

227. **Lee HS, Zhang Y.** 2012. BSP-SLIM: a blind low-resolution ligand-protein docking approach using predicted protein structures. Proteins **80:**93-110.

228. **Elsinghorst EA.** 1994. Measurement of invasion by gentamicin resistance. Methods Enzymol. **236:**405-420.

229. **Klimpel GR, Shaban R, Niesel DW.** 1990. Bacteria-infected fibroblasts have enhanced susceptibility to the cytotoxic action of tumor necrosis factor. J. Immunol. **145:**711-717.

230. **Kapasi K, Inman RD.** 1992. HLA-B27 expression modulates gram-negative bacterial invasion into transfected L cells. J. Immunol. **148:**3554-3559.

231. **Okamura N, Nakaya R.** 1977. Rough mutant of *Shigella flexneri* 2a that penetrates tissue culture cells but does not evoke keratoconjunctivitis in guinea pigs. Infect. Immun. **17:**4-8.

232. **Sandlin RC, Goldberg MB, Maurelli AT.** 1996. Effect of O side-chain length and composition on the virulence of *Shigella flexneri* 2a. Mol Microbiol. **22:**63-73.

233. **Allaoui A, Menard R, Sansonetti PJ, Parsot C.** 1993. Characterization of the *Shigella flexneri ipgD* and *ipgF* genes, which are located in the proximal part of the *mxi* locus. Infect. Immun. **61:**1707-1714.

234. **Pendaries C, Tronchere H, Arbibe L, Mounier J, Gozani O, Cantley L, Fry MJ, Gaits-Iacovoni F, Sansonetti PJ, Payrastre B.** 2006. PtdIns5P activates the host cell PI3-kinase/Akt pathway during *Shigella flexneri* infection. EMBO J. **25:**1024-1034.

235. **Niebuhr K, Giuriato S, Pedron T, Philpott DJ, Gaits F, Sable J, Sheetz MP, Parsot C, Sansonetti PJ, Payrastre B.** 2002. Conversion of PtdIns(4,5)P(2) into PtdIns(5)P by the *Shigella flexneri* effector IpgD reorganizes host cell morphology. EMBO J. **21:**5069-5078.

236. **Nakayama GR, Caton MC, Nova MP, Parandoosh Z.** 1997. Assessment of the Alamar Blue assay for cellular growth and viability *in vitro*. J. Immunol. Methods. **204:**205-208.

237. **Manzanera M, Marques S, Ramos JL.** 2000. Mutational analysis of the highly conserved C-terminal residues of the XylS protein, a member of the AraC family of transcriptional regulators. FEBS Lett. **476:**312-317.

238. **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.

239. **Swint-Kruse L, Brown CS.** 2005. Resmap: automated representation of macromolecular interfaces as two-dimensional networks. Bioinformatics **21:**3327-3328.

240. **Singh J, Thornton JM.** 1992. Atlas of Protein Side-Chain Interactions, vol. 1. IRL Press, Oxford.

241. **Maloy SR, Stewart VJ, Taylor RK.** 1996. Genetic analysis of pathogenic bacteria. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.