Elucidating Protein-DNA Recognition by the Bacterial Transcription Factors MarA and VirF by Homology Modeling and DNA-Binding Studies of Wild-type and Mutant MarA and VirF

by

Nicholas James Ragazzone

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Medicinal Chemistry)
in the University of Michigan
2022

Doctoral Committee:

Professor George A. Garcia, Chair
Professor Amanda Garner
Professor Nouri Neamati
Professor David Sherman

Nicholas Ragazzone

nragazzo@umich.edu

ORCID iD: 0000-0002-3191-0301

Dedication

This dissertation is dedicated to my mother, Sarah. She always had faith in me, even when I struggled to find faith in myself.

Acknowledgements

I would not be where I am today without the support of several people. First, I want to express my sincere and heartfelt gratitude for my advisor, Dr. George Garcia. You have provided me with guidance and support through one of the greatest challenges of my life. I will never forget my time in the lab, and I cannot thank you enough for helping me develop into a well-rounded scientist and person. I would also like to thank my dissertation committee for their helpful advice through the constant challenges of my research: Dr. Amanda Garner, Dr. Nouri Neamati, and Dr. David Sherman.

I would also like to show appreciation to my former lab members for their patience towards me as I tried to figure out what I was doing in the lab. Dr. Anthony Emanuele for being another mentor and helping guide me as I took over his work, I should have listened when you told me how hard this project was going to be. Dr. Maxwell Stefan and Dr. Nathan Scharf for helping me troubleshoot different problems as well as teaching me what they knew. I want to thank my current lab members as well. Shireen Ashkar, Katie Guild, and Glory Velazquez for all their help as well as their company over these years. It's been a pleasure working with each of you. I also want to thank Garrett Dow. You know as well as I do how difficult this work has been, and I appreciate being able to bounce ideas off each other. I also appreciate having someone to talk to about music, as well as playing IM sports together. It was always a great way to blow off steam after a long day in lab.

Completing my PhD would not have been possible without the love and support of all my friends and family. My friends from before graduate school, I am so grateful that we have all

been able to stay close over such a long time and distance apart. My cohort, who have all gone on to do great things since leaving the program. I could not have asked for a better network of people to help each other through such a difficult period. I want to give thanks to my roommates throughout graduate school. When I moved into 903, I never would have guessed I was about to make friends for life. You all were such a big part of keeping me sane and happy as we worked through our PhDs. I want to thank all my siblings: Mike, Mario, Desirae, and Christian. Moving so far away from all of you was such a difficult decision, but I cherish all your love and support for me through this time. From the bottom of my heart, I want to thank my parents. My dad, Mario, for helping me move to Michigan all those years ago and my mom, Sarah, who was so proud of everything I did. I love you and would trade anything to have you here with me as I finish my PhD. Lastly, I want to thank my wonderful partner, Morgan. You have been my rock as I approached the finish line. There were so many days where I would come home beaten down and discouraged and you have always been able to lift me up and convince me to keep going. I can't express how thankful I am for your love and patience over these last several years.

Table of Contents

List of Tables

List of Figures

Abstract

Shigella flexneri, a gram-negative pathogen, is the main cause of bacterial dysentery in humans. Infections by *Shigella*, known as shigellosis, lead to approximately 200,000 deaths globally each year. Current treatments include ciprofloxacin and azithromycin, but the resistance rates to these antibiotics have risen significantly worldwide. This highlights a critical need for novel treatments for bacterial infections. One such approach that may potentially circumvent the raise of resistance is to target bacterial tools for infection, rather than bacterial viability. This approach is known as virulence inhibition or antivirulence.

Shigella relies on various virulence factors that are essential to macrophage apoptosis and escape, intestinal epithelial cell invasion and cell-to-cell spread. These processes rely on a main transcriptional regulator, VirF, to activate transcription of the virulence genes *virB* and *icsA*. While several AraC proteins have been studied, almost none have reported using native VirF, and the three-dimensional structure is yet to be solved. We hypothesize that VirF would make an ideal antivirulence target, and here, we set out to better understand how VirF interacts with DNA in order to gain insight as to how this interaction can be probed as a target for inhibition.

To work with VirF *in vitro*, researchers have relied on a large maltose binding protein tag that solubilizes the protein. This leaves the question of whether any of the results from these studies would be different with the native form of the protein. To address this, we attempted to optimize the expression and purification of a N-terminal histidine-tagged VirF and a truncated form of the VirF DNA binding domain (DBD). We further characterized the VirF DBD using the structures of two *E. coli* VirF homologs, GadX and MarA•*marRAB*, to generate homology

models of the VirF DNA-binding domain in free and DNA-bound conformations. We conducted an alanine scan of seven residues in MarA and VirF that make base-specific interactions to identify residues important for binding to the *marRAB* and *virB* promoters, respectively. On helix 3 of MarA, we found that mutating W42, R46, and R96 significantly reduced the ability of MarA to bind the *marRAB* promoter. Furthermore, when mutating the corresponding residues on the VirF binding site to alanine, we found that each mutant displayed weaker binding to the *virB* promoter relative to WT. This indicated that these residues are important for binding, supporting our homology model. We continued to probe the VirF•*virB* interaction by developing chimeric proteins of MarA and VirF, hoping to induce binding to *virB* with the goal of using the chimeric protein as a model to study the VirF DBD inhibitor, 19615. Unfortunately, we were unsuccessful at mutating MarA in such a way to coerce binding to the *virB* promoter. Further efforts are needed to design and purify mutants of MarA that may be able to recognize *virB*.

Chapter 1 Introduction

Antimicrobial Resistance and Targeting Virulence

The emergence and spread of antimicrobial resistance (AMR) has been a worldwide concern for several years. AMR infections take 50,000 lives between Europe and the United States and an estimated 700,000 lives globally each year.[1,2] Between 1960 and 2000, the pharmaceutical industry released several antimicrobials to combat the rise of resistant infections, but resistance has been identified to virtually every antimicrobial that has been developed.[3] Over the past 30 years, only one new class of antimicrobials has been discovered,[4,5] while the number of drug resistant pathogens has continued to increase at an alarming rate.[6,7] To exacerbate the issue, antimicrobials are widely misused and overused across the globe despite studies demonstrating that antimicrobial consumption and the emergence of resistant strains share a direct relationship.[8] The inappropriate prescribing of antibiotics contributes greatly to the rise of drug resistant bacterial infections and studies have shown that 30% to 60% of the time, antibiotics prescribed in U.S. intensive care units are unnecessary, inappropriate or suboptimal.[9] Their over-the-counter availability makes access easy as well, leading to the misuse by the community in addition to healthcare professionals. In addition, patient compliance and unregulated supply chains, especially in developing nations, contributes greatly to the problem. If left uncontrolled, AMR may have substantial human and economic cost. Medical procedures that are common and safe today will potentially be high risk for infection with lower success rates and can lead to lengthier and more expensive hospital stays. It has been projected that by 2050, 10 million people would die each year and there would be a global decrease of 2-3.5% in

Gross Domestic Product (GDP) due to antimicrobial resistance.[1,10,11] The World Health Organization (WHO) has responded by placing AMR on its list of the top 10 Global Health Issues in 2021.[12] This effort will help bring resources to combat antimicrobial resistance but the need for new classes of antimicrobials remains urgent.

Traditional bactericidal or bacteriostatic antibiotics are designed to kill or impair the growth of their targets respectively. They are effective by targeting functions that are important for bacterial growth such as cell wall synthesis (penicillins), DNA replication (quinolones), RNA transcription (rifamycins), and protein synthesis (tetracyclines, macrolides).[13] While effective targets, these pathways are shared broadly among diverse bacteria, contributing to a strong selective pressure that fosters the development of antibiotic resistant strains.[14] For example, any population of bacteria will naturally contain a small population that will harbor resistance to an antibiotic due to natural bacterial mutation rates. Once most of the non-resistant bacteria are killed, the small, resistant population can quickly grow and become the dominate population. One relatively novel approach is to disarm pathogens in the host by targeting bacterial virulence. The overarching strategy aims to inhibit specific mechanisms that promote infection and are crucial for pathogen persistence or factors that cause symptoms.[15,16] This approach offers several potential benefits when compared to traditional methods, including: increased diversity of pharmacological targets and mechanisms of action for novel inhibitors, reduced chances of resistant development due to lower selective pressure on the bacterial life cycle, lower risk related to environmental exposure due to the fact that many virulence targets are only expressed inside of the host, and the potential to preserve the non-pathogenic, natural gut microbiota and avoid infection by opportunistic pathogens such as *Clostridium difficile*.[17]

There are several approaches to antivirulence based drug discovery. One approach is to neutralize bacterial toxins and inhibit toxin transcription. The cholera toxin in *Vibrio cholerae* is key in the severe symptoms caused by the infection and neutralizing its activity would theoretically alleviate the symptoms.[18,19] Another strategy is blocking bacterial adherence by targeting particular surface receptors recognized by bacteria or bacterial pili. For example, uropathogenic *E. coli* rely on type 1 and P pili to attach and invade their target cells and methods to interrupt pili assembly and block their adhesive properties are being investigated.[20,21] In addition, bacterial communications can be targeted by interrupting intercellular chemical signaling. This affects the ability of the bacteria to track their cellular density (i.e., quorum sensing) as well as other community behavior processes and virulence, ultimately making the biofilms more susceptible to host immune responses and antibacterial intervention.[22] Specialized bacterial secretion systems, like the type III secretion system, are used for injecting effector proteins into the host cell. These effectors imitate host protein function and significantly affect signaling pathways, contributing greatly to the progression of the illness. These secretion systems make attractive targets and provide multiple points of intervention such as base or syringe assembly, host interaction, and effector secretion.[23–26] Lastly, strategies to develop inhibitors of transcription factors that regulate virulence gene expression are being investigated as an antivirulence therapy. Inhibiting the expression of these factors can ameliorate the infection by interfering with the bacteria's ability to spread and establish infection and manipulate host immune responses.[27,28]

Shigella and VirF

One pathogen labeled as a "Serious Threat" by the Center for Diseases Control (CDC) is the enteropathogen, *Shigella*.[29] *Shigella* spp., the causative organism of the diarrheal disease Shigellosis, lead to nearly 270 million cases and over 200,000 deaths globally each year.[30–32] Currently, the genus *Shigella* is divided into four species (also called subgroups): *S. dysenteriae*, *S. flexneri*, *S. boydii*, and *S. sonnei*. While *S. sonnei* may be the most prevalent in Europe and North America (80% of infections in these regions), *S. flexneri* is the major cause of dysentery in low-income regions such as South Asia and Sub-Saharan Africa (up to 62% of all *Shigella* spp. Infections).[31] Individuals suffering from Shigellosis experience fever, abdominal pain, dehydration, and diarrhea that is often bloody and mucoidal. *Shigella* is transmitted from host to host through fecal-to-oral contact, which correlates to the fact that the majority of cases of shigellosis occur in nations that lack access to proper sanitation, clean drinking water, and adequate healthcare.[33,34] Additionally, most bacterial pathogens require millions of bacteria to establish an infection, where shigellosis requires as little as 10 viable organisms.[35] Most domestic cases occur in day-care centers or assisted living facilities, although the total number of cases in the United States are only about 450,000 annually. Still, that results in an estimated $93 million in medical costs, increasing year over year as the number of emerging drug-resistant *Shigella* clinical strains increases.[30,36] Worldwide, populations most at risk are children under the age of five (69% of all cases and 61% of all deaths), men who have sex with men, immunocompromised individuals, and travelers to regions with inadequate sanitation.[29,37,38]

For acute *Shigella* infections, the recommended first-line treatment is ciprofloxacin, followed by third generation cephalosporins and azithromycin.[37,39] Recently, ciprofloxacin resistance in *Shigella* spp. has been reported to be rising in several countries with a median

resistance of 19.4%.[37] Fluoroquinolones target DNA gyrase, which is essential of DNA replication and transcription, but mutations to the gyrA gene reduce efficacy of these drugs in *Shigella* and other enteropathogens. Recent *S. sonnei* isolates in the USA, Vietnam and India have been reported to have complete ciprofloxacin resistance (MIC ≥ 4 mg/L).[40–42] For several countries, *Shigella* clinical isolates have been reported to be resistant to cephalosporins such as ceftriaxone in an average of 18.5% of cases.[37] This observed resistance is likely due to the production of extended-spectrum β-lactamases (ESBLs), which can inactivate the antibiotics through hydrolysis of the β-lactam ring.[39]

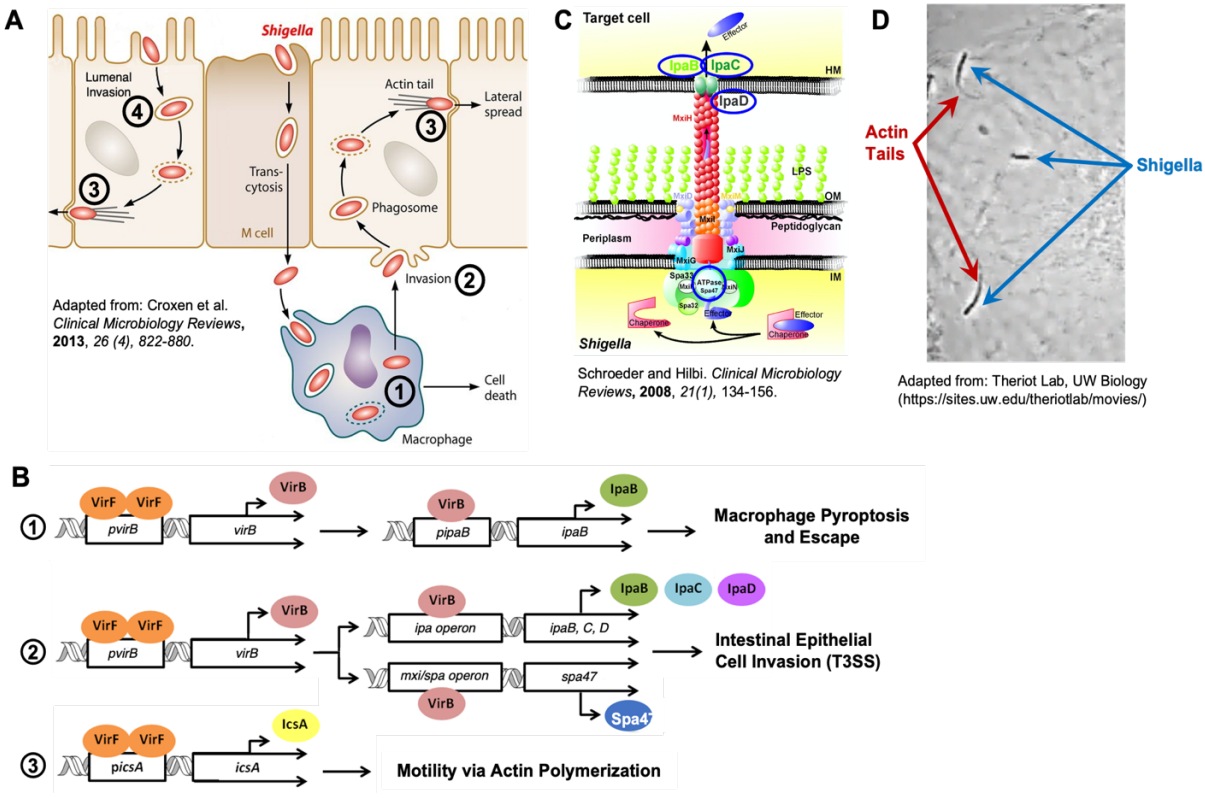


Figure 1-1: ***Shigella Pathogenesis Pathway****. A): Shigella escape from the macrophages (1), invasion of epithelial cells (2), and lateral spread (3). B): Primary virulence factor induction by VirF. C): Type III secretion system structure. D): Lateral spread via actin polymerization facilitated by IcsA.*

The molecular mechanism used by *Shigella* to invade and replicate in the host's intestinal epithelium has been studied extensively and is summarized in **Figure 1-1**.[31,43–48] To establish an

infection, *Shigella* relies on the coordinated expression of virulence genes spread across pathogenicity islands on the chromosome, and on the large 230 kbp virulence plasmid (pINV) **(Figure 1-2)**.[49] Once ingested, *Shigella* can resist the high acidity of the stomach to reach the colon, contributing to the extremely low number of bacteria required to establish an infection. This acid resistance is largely due to the growth phase-dependent sigma factor σ^{38}, coded by *rpoS*.[50] *Shigella* is able to enter the colonic epithelium through two pathways. The first is mediated by host microfold-cells (M-cells). *Shigella* enters the M-cells though induced membrane ruffling by the secreted effector protein IpgB1, beginning transcytosis to the submucosa of the gastrointestinal tract.[51] Here, the bacteria is endocytosed by resident macrophages but escape degradation by inducing immediate macrophage pyroptosis. This is achieved through a caspase 1-dependent pathway, releasing proinflammatory cytokines (IL-1 and IL-8) that destabilize the junctions between epithelial cells by recruiting polymorphonuclear (PMN) leukocytes.[52] This destabilization facilitates *Shigella* invasion, bypassing the M-cells, leading to the second pathway *Shigella* uses to enter the colonic epithelium. Once in contact with the basolateral membrane of the intestinal epithelial cells, the bacteria express the type III secretion system (T3SS). This hollow, needle-like structure is composed of several proteins from the *mxi-spa* locus of pINV and is important in releasing effector proteins into the epithelial cell during invasion. IpaB at the tip of the T3SS binds to host CD44 to stabilize the bacterium as IpaB, IpaC, and IpaD, released by the T3SS, interact with host receptors such as $\alpha_5\beta_1$ integrin to begin rearrangement of the host cytoskeleton, promoting uptake of the bacterium.[53–55] Effector proteins IpaB, IpaC, IpaD, and IpaH facilitate the lysis of the phagosome, releasing *Shigella* into the epithelial cytoplasm.[31,56] Here, *Shigella* can replicate and survive by preventing epithelial cell death by releasing IpgD and VirA, which contribute to augmenting pro-survival signaling.[55]

Shigella lack any flagella for movement. Instead, IcsA mediates the recruitment of the host's N-WASP and ARP2/3 at one pole of the bacterium to create a complex that acts as a nucleation site for host actin polymerization to propel itself forward.[57–59] This method of motility allows *Shigella* to make protrusions at tricellular tight junctions to become endocytosed by the neighboring epithelial cell to start the escape, replication and spread cycle over.

On the pINV, a 31 kb region known as the "entry region" contains most of the genes that are required for *Shigella* pathogenesis. The 34 genes located here are split into an *ipa* cluster, which largely encode for the Ipa effector proteins, and a *mxi-spa* cluster, which encode for the MxiE regulator, and the T3SS structural machinery. *Ipg* genes are scattered between the two clusters.[60,61] Genes located outside of the entry region that are critical for invasion include *icsA*, *virA*, *virB*, and *virF*. As mentioned previously, IcsA is important for *Shigella* motility, and VirA plays a role in interfering with cell autophagy and disrupting the entry vacuole.[43,44,62] VirB is a transcriptional regulator that activates the expression of the *mxi-spa* genes that form the T3SS and the early Ipa effectors. Additionally, VirB turns on the

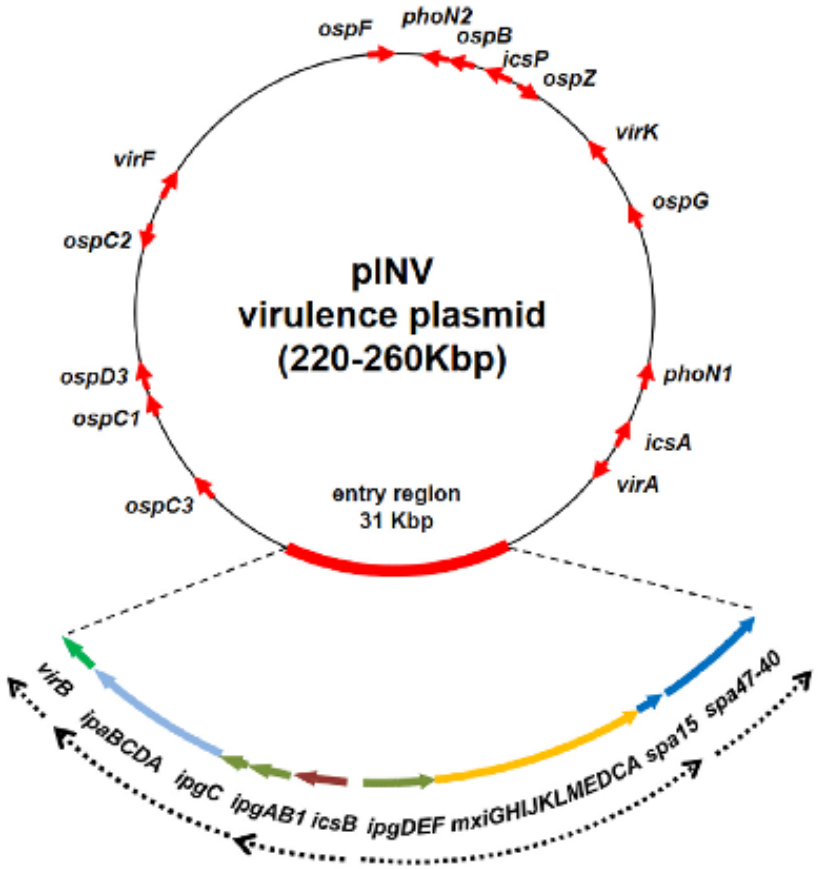


*Figure 1-2:**Plasmid Map of pINV**. Image was taken from Pasqua et al.*

expression of the MxiE regulator which in turn is responsible for activation of the late effectors in the invasion process. The *virF* gene encodes for the transcriptional regulator VirF, which is responsible for activating *icsA* and *virB*, putting VirF at the top of the regulatory cascade of *Shigella* virulence gene expression.[43,63,64]

VirF, a member of the AraC family of transcriptional regulators, is crucial in the initiation of the *Shigella* pathogenesis pathway. Tobe *et al.*(1993)[65] first demonstrated the ability of VirF to activate *virB*. The authors were able to use deletion analysis of the *virB* promoter to identify an important region 110 bp upstream of the *virB* transcription start site. This region correlated with binding by the MalE-VirF fusion protein as well binding by the transcription silencer, H-NS. The overlapping binding sites of VirF and H-NS support the theory of temperature dependent regulation. It is thought that an increase in temperature from 30°C to 37°C can induce a change in DNA supercoiling that allows for favorable conditions for VirF binding and disrupts the H-NS • DNA complex that represses transcription of *Shigella* virulence genes.[65] In the case of *icsA*, transcription is repressed by the binding of H-NS to three unique sites as well as by the small antisense RNA, RnaG.[66,67] VirF is able to counteract this repression, promoting *icsA* expression through two pathways: by directly competing with H-NS binding to switch to an active state, and by repressing the expression of RnaG.[68]

DNA Binding Domain

AraC proteins are characterized by their conserved C-terminal helix-turn-helix (HTH) containing DNA binding domain **(Figure 1-3)**.[69] They also typically contain a nonconserved N-terminal domain that is often responsible for dimerization or chemical signaling.[70] The mechanism by which this class of proteins binds DNA has been studied through the two co-

crystal structures of MarA (PDB:1BL0) and Rob (PDB:1D5Y) each bound to their target DNA.[71,72] These two structures suggest two different mechanisms of binding by the C-terminal DNA binding region. In the case of MarA, the conserved HTH domains bind with each helix inserted into the DNA major groove.[73] In the crystal structure of Rob, the protein is bound with only one helix inserted into the major groove while the other end of the protein interacts with the surface of the DNA double helix. Recent studies have questioned the validity of this Rob structure, using cryogenic electron microscopy (cryo-EM) to obtain a density map and structural model of Rob in a transcription activation complex.[74,75] Similar to the MarA crystal structure, the cryo-EM structure of Rob shows the HTH motifs interacting with DNA in the major grooves, resulting in a 35° bend in the DNA.[71,74] Porter and Dorman were able to provide insight to the structure and function of VirF through full gene random and site-directed mutagenesis.[76] Changes to the first and second HTH regions inactivated VirF in a β-galactosidase assay, specifically I180N, V191A, K193A, and E196K. When comparing the outcome of these mutations with the alanine scan on MarA[77], trends can be observed in positions that are important for both proteins. For example, K193 in VirF likely makes major groove contacts similarly to R46 in MarA, explaining the severe change in phenotype when this residue is mutated.[76] This data, along with similar results seen with another AraC regulator, PerA, provide evidence that VirF interacts with its target DNA with both HTH regions of the DNA binding domain.[64]

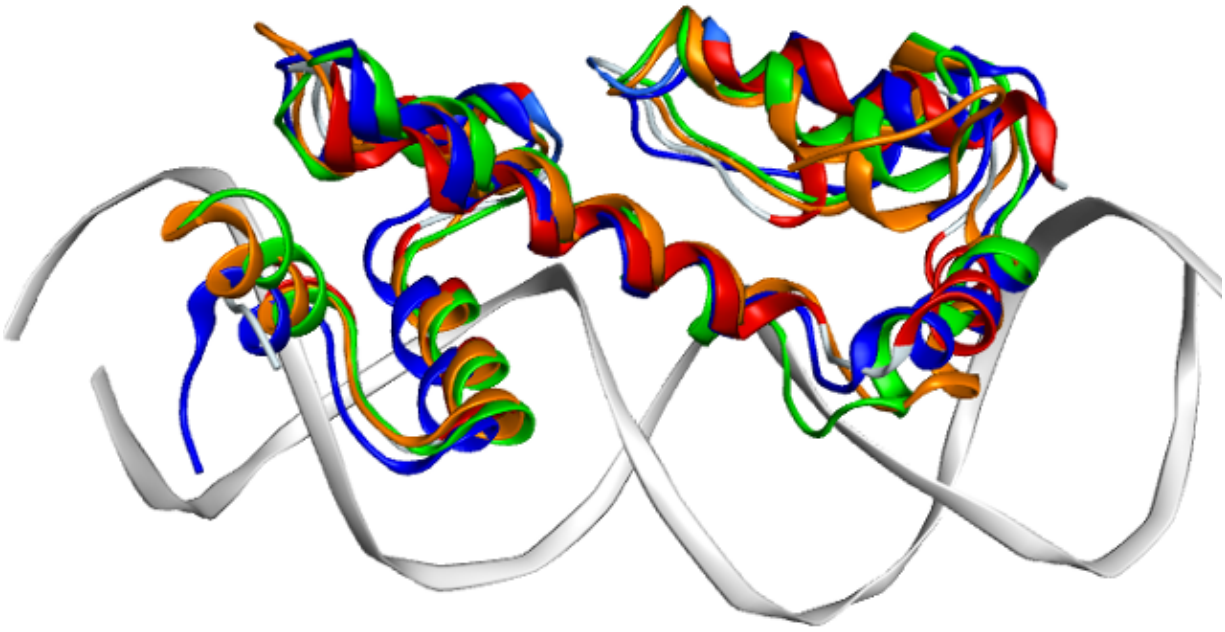


Figure 1-3: ***Overlay of the DNA Binding Domains of Four AraC Family Members.*** *Structures included are MarA (1BL0), AraC (2K9S), GadX (3MKL), and the DNA binding domain of ToxT (5SUW).*

In this research project, we focus on studying the VirF DNA binding domain (DBD). In **Chapter 2**, I will discuss attempts to express and purify an active, truncated form of the VirF DBD. Next, in **Chapter 3**, we used two *E. coli* homologs, GadX and MarA, to generate homology models of the VirF DBD. These models guided our alanine scan mutagenesis of MarA and MalE-VirF, and *in vitro* DNA binding assays identified resides that were important in both proteins' capabilities to bind DNA. In **Chapter 4**, we construct domain swaps of the different MarA and VirF binding regions to evaluate whether the proteins can be made to recognize their non-native DNA ligand. Lastly, **Chapter 5** will briefly discuss attempts to run *Shigella* invasion and plaque formation assays on Caco-2 human epithelial cells as part of a 1.7 M compound phenotypic, high-throughput screen conducted in collaboration with GSK.

References

(1) O'Neill, J. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations The Review on Antimicrobial Resistance Chaired. **2014**, No. December.

(2) Pokharel, S.; Adhikari, B. Antimicrobial Resistance and over the Counter Use of Drugs in Nepal. *J. Glob. Health* **2020**, *10* (1), 1–4. https://doi.org/10.7189/JOGH.10.010360.

(3) Ventola, C. L. The Antibiotic Resistance Crisis: Part 1: Causes and Threats. *Pharm. Ther.* **2015**, *40* (4), 277–283. https://doi.org/Article.

(4) Chahine, E. B.; Karaoui, L. R.; Mansour, H. Bedaquiline: A Novel Diarylquinoline for Multidrug-Resistant Tuberculosis. *Ann. Pharmacother.* **2014**, *48* (1), 107–115. https://doi.org/10.1177/1060028013504087.

(5) Leibert, E.; Danckers, M.; Rom, W. N. New Drugs to Treat Multidrug-Resistant Tuberculosis: The Case for Bedaquiline. *Ther. Clin. Risk Manag.* **2014**, *10* (1), 597–602. https://doi.org/10.2147/TCRM.S37743.

(6) Aminov, R. History of Antimicrobial Drug Discovery: Major Classes and Health Impact. *Biochem. Pharmacol.* **2017**, *133*, 4–19. https://doi.org/10.1016/j.bcp.2016.10.001.

(7) Durand, G. A.; Raoult, D.; Dubourg, G. Antibiotic Discovery: History, Methods and Perspectives. *Int. J. Antimicrob. Agents* **2019**, *53* (4), 371–382. https://doi.org/10.1016/j.ijantimicag.2018.11.010.

(8) The Antibiotic Alarm. *Nature* **2011**, No. 5637, 2013.

(9) Chiotos, K.; Tamma, P. D.; Gerber, J. S. Antibiotic Stewardship in the Intensive Care Unit: Challenges and Opportunities. *Infect. Control Hosp. Epidemiol.* **2019**, *40* (6), 693–698. https://doi.org/10.1017/ice.2019.74.

(10) Ayukekbong, J. A.; Ntemgwa, M.; Atabe, A. N. The Threat of Antimicrobial Resistance in Developing Countries: Causes and Control Strategies. *Antimicrob. Resist. Infect. Control* **2017**, *6* (1), 1–8. https://doi.org/10.1186/s13756-017-0208-x.

(11) Mulani, M. S.; Kamble, E. E.; Kumkar, S. N.; Tawre, M. S.; Pardesi, K. R. Emerging Strategies to Combat ESKAPE Pathogens in the Era of Antimicrobial Resistance: A Review. *Front. Microbiol.* **2019**, *10* (APR). https://doi.org/10.3389/fmicb.2019.00539.

(12) World Health Organization (WHO). Antimicrobial Resistance https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance.

(13) Clatworthy, A. E.; Pierson, E.; Hung, D. T. Targeting Virulence: A New Paradigm for Antimicrobial Therapy. *Nat. Chem. Biol.* **2007**, *3* (9), 541–548. https://doi.org/10.1038/nchembio.2007.24.

(14) DIckey, S. W.; Cheung, G. Y. C.; Otto, M. Different Drugs for Bad Bugs: Antivirulence Strategies in the Age of Antibiotic Resistance. *Nat. Rev. Drug Discov.* **2017**, *16* (7), 457–471. https://doi.org/10.1038/nrd.2017.23.

(15) Rasko, D. a.; Sperandio, V. Anti-Virulence Strategies to Combat Bacteria-Mediated Disease. *Nat. Rev. Drug Discov.* **2010**, *9* (2), 117–128. https://doi.org/10.1038/nrd3013.

(16) Cegelski, L.; Marshall, G. R.; Eldridge, G. R.; Hultgren, S. J. The Biology and Future Prospects of Antivirulence Therapies. *Nat. Rev. Microbiol.* **2008**, *6* (1), 17–27. https://doi.org/10.1038/nrmicro1818.

(17) Heras, B.; Scanlon, M. J.; Martin, J. L. Targeting Virulence Not Viability in the Search for Future Antibacterials. *Br. J. Clin. Pharmacol.* **2015**, *79* (2), 208–215. https://doi.org/10.1111/bcp.12356.

(18) Hung, D. T.; Shakhnovich, E. A.; Pierson, E.; Mekalanos, J. J. Small-Molecule Inhibitor of Vibrio Cholerae Virulence and Intestinal Colonization. *Science (80-.).* **2005**, *310* (5748), 670–674. https://doi.org/10.1126/science.1116739.

(19) Spangler, B. D. Structure and Function of Cholera Toxin and the Related Escherichia Coli Heat-Labile Enterotoxin. *Microbiol. Rev.* **1992**, *56* (4), 622–647. https://doi.org/10.1128/mr.56.4.622-647.1992.

(20) Bouckaert, J.; Berglund, J.; Schembri, M.; De Genst, E.; Cools, L.; Wuhrer, M.; Hung, C. S.; Pinkner, J.; Slättegård, R.; Zavialov, A.; et al. Receptor Binding Studies Disclose a Novel Class of High-Affinity Inhibitors of the Escherichia Coli FimH Adhesin. *Mol. Microbiol.* **2005**, *55* (2), 441–455. https://doi.org/10.1111/j.1365-2958.2004.04415.x.

(21) Sauer, F. G.; Remaut, H.; Hultgren, S. J.; Waksman, G. Fiber Assembly by the Chaperone-Usher Pathway. *Biochim. Biophys. Acta - Mol. Cell Res.* **2004**, *1694* (1-3 SPEC.ISS.), 259–267. https://doi.org/10.1016/j.bbamcr.2004.02.010.

(22) Hentzer, M.; Wu, H.; Andersen, J. B.; Riedel, K.; Rasmussen, T. B.; Bagge, N. Attenuation of Pseudomonas Aeruginosa Virulence by Quorum-Sensing Inhibitors. *Embo J.* **2003**, *22* (15), 3803.

(23) Muschiol, S.; Bailey, L.; Gylfe, Å.; Sundin, C.; Hultenby, K.; Bergström, S.; Elofsson, M.; Wolf-Watz, H.; Normark, S.; Henriques-Normark, B. A Small-Molecule Inhibitor of Type III Secretion Inhibits Different Stages of the Infectious Cycle of Chlamydia Trachomatis. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (39), 14566–14571. https://doi.org/10.1073/pnas.0606412103.

(24) Kauppi, A. M.; Nordfeith, R.; Uvell, H.; Wolf-Watz, H.; Elofsson, M. Targeting Bacterial Virulence: Inhibitors of Type III Secretion in Yersinia. *Chem. Biol.* **2003**, *10*, 241–249. https://doi.org/10.1016/S.

(25) Veenendaal, A. K. J.; Sundin, C.; Blocker, A. J. Small-Molecule Type III Secretion System Inhibitors Block Assembly of the Shigella Type III Secreton. *J. Bacteriol.* **2009**, *191* (2), 563–570. https://doi.org/10.1128/JB.01004-08.

(26) Gauthier, A.; Robertson, M. L.; Lowden, M.; Ibarra, J. A.; Puente, J. L.; Finlay, B. B. Transcriptional Inhibitor of Virulence Factors in Enteropathogenic Escherichia Coli. *Antimicrob. Agents Chemother.* **2005**, *49* (10), 4101–4109. https://doi.org/10.1128/AAC.49.10.4101-4109.2005.

(27) Gordon, C. P.; Williams, P.; Chan, W. C. Attenuating Staphylococcus Aureus Virulence Gene Regulation: A Medicinal Chemistry Perspective. *J. Med. Chem.* **2013**, *56* (4), 1389–1404. https://doi.org/10.1021/jm3014635.

(28) Koppolu, V.; Osaka, I.; Skredenske, J. M.; Kettle, B.; Hefty, P. S.; Li, J.; Egan, S. M. Small-Molecule Inhibitor of the Shigella Flexneri Master Virulence Regulator VirF. *Infect. Immun.* **2013**, *81* (11), 4200–4207. https://doi.org/10.1128/IAI.00919-13.

(29) Centers for Disease Control and Prevention. Antibiotic Resistance Threats in the United States. *U.S. Dep. Heal. Hum. Servies* **2019**, 1–113. https://doi.org/10.15620/cdc:82532.

(30) Puzari, M.; Sharma, M.; Chetia, P. Emergence of Antibiotic Resistant Shigella Species: A Matter of Concern. *J. Infect. Public Health* **2018**, *11* (4), 451–454. https://doi.org/10.1016/j.jiph.2017.09.025.

(31) The, H. C.; Thanh, D. P.; Holt, K. E.; Thomson, N. R. The Genomic Signatures of Shigella Evolution, Adaptation and Geographical Spread. *Nat. Publ. Gr.* **2016**, *14* (4), 235–250. https://doi.org/10.1038/nrmicro.2016.10.

(32) Khalil, I. A.; Troeger, C.; Blacker, B. F.; Rao, P. C.; Brown, A.; Atherly, D. E.; Brewer, T. G.; Engmann, C. M.; Houpt, E. R.; Kang, G.; et al. Morbidity and Mortality Due to Shigella and Enterotoxigenic Escherichia Coli Diarrhoea: The Global Burden of Disease Study 1990–2016. *Lancet Infect. Dis.* **2018**, *18* (11), 1229–1240. https://doi.org/10.1016/S1473-3099(18)30475-4.

(33) Stevens, P. Diseases of Poverty and the 10/90 Gap. In *Fighting the Diseases of Poverty*; 2004.

(34) Médecins Sans Frontières. *Fatal Imbalance: The Crisis in Research and Development for Drugs for Neglected Diseases*; 2001.

(35) Formal, S. B. Inoculum Size in Shigellosis and Implications For Expected Mode of Transmission. *J. Infect. Dis.* **1989**, *159* (6), 1126–1128. https://doi.org/10.1093/infdis/159.6.1126.

(36) Frieden, T. Antibiotic Resistance Threats. *Cdc* **2013**, 22–50. https://doi.org/CS239559-B.

(37) Antimicrobial Resistance Division; Global Antimicrobial Resistance Surveillance System; Global Antimicrobial Resistance Surveillance System (GLASS); SurveillancePrevention&Control; Agnew, E.; Dolecek, C.; Hasan, R.; Lahra, M.; Merk, H.; Perovic, O.; et al. *Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report 2021*; 2021.

(38) Klontz, K. C.; Singh, N. Treatment of Drug-Resistant *Shigella* Infections. *Expert Rev. Anti. Infect. Ther.* **2015**, *13* (1), 69–80. https://doi.org/10.1586/14787210.2015.983902.

(39) Williams, P.; Berkley, J. A. Dysentery (Shigellosis) Current WHO Guidelines and the WHO Essential Medicine List for Children http://www.who.int/selection_medicines/committees/expert/21/applications/s6_paed_antibiotics_appendix5_dysentery.pdf.

(40) Bowen, A.; Hurd, J.; Hoover, C.; Khachadourian, Y.; Traphagen, E.; Harvey, E.; Libby, T.; Ehlers, S.; Ongpin, M.; Norton, J. C.; et al. Importation and Domestic Transmission of Shigella Sonnei Resistant to Ciprofloxacin - United States, May 2014-February 2015. *MMWR. Morb. Mortal. Wkly. Rep.* **2015**, *64* (12), 318–320.

(41) Kim, J. S.; Kim, J. J.; Kim, S. J.; Jeon, S. E.; Seo, K. Y.; Choi, J. K.; Kim, N. O.; Hong, S.; Chung, G. T.; Yoo, C. K.; et al. Outbreak of Ciprofloxacin-Resistant Shigella Sonnei Associated with Travel to Vietnam, Republic of Korea. *Emerg. Infect. Dis.* **2015**, *21* (7), 1247–1250. https://doi.org/10.3201/eid2107.150363.

(42) De Lappe, N.; O'Connor, J.; Garvey, P.; McKeown, P.; Cormican, M. Ciprofloxacin-Resistant Shigella Sonnei Associated with Travel to India. *Emerg. Infect. Dis.* **2015**, *21* (5), 894–896. https://doi.org/10.3201/eid2105.141184.

(43) Di Martino, M. L.; Falconi, M.; Micheli, G.; Colonna, B.; Prosseda, G. The Multifaceted Activity of the VirF Regulatory Protein in the Shigella Lifestyle. *Front. Mol. Biosci.* **2016**, *3* (SEP), 1–11. https://doi.org/10.3389/fmolb.2016.00061.

(44) Mattock, E.; Blocker, A. J. How Do the Virulence Factors of Shigella Work Together to Cause Disease? *Front. Cell. Infect. Microbiol.* **2017**, *7* (MAR), 1–24. https://doi.org/10.3389/fcimb.2017.00064.

(45) Zychlinsky, A.; Sansonetti, P. J. Shigella Host Cell Invasion: Signficance in Pathogenesis. *Adv. Cell. Mol. Biol. Membr. Organelles* **1999**, *6*, 181–200. https://doi.org/10.1016/S1874-5172(99)80012-1.

(46) Croxen, M. A.; Law, R. J.; Scholz, R.; Keeney, K. M.; Wlodarska, M.; Finlay, B. B. Recent Advances in Understanding Enteric Pathogenic Escherichia Coli. *Clin. Microbiol. Rev.* **2013**, *26* (4), 822–880. https://doi.org/10.1128/CMR.00022-13.

(47) Schroeder, G. N.; Hilbi, H. Molecular Pathogenesis of Shigella Spp.: Controlling Host Cell Signaling, Invasion, and Death by Type III Secretion. *Clin. Microbiol. Rev.* **2008**, *21* (1), 134–156. https://doi.org/10.1128/CMR.00032-07.

(48) Theriot, J.; Portnoy, D. Theriot Lab at the University of Washington: Publications and Movies https://sites.uw.edu/theriotlab/movies/.

(49) Pasqua, M.; Michelacci, V.; Di Martino, M. L.; Tozzoli, R.; Grossi, M.; Colonna, B.; Morabito, S.; Prosseda, G. The Intriguing Evolutionary Journey of Enteroinvasive E. Coli (EIEC) toward Pathogenicity. *Front. Microbiol.* **2017**, *8* (DEC). https://doi.org/10.3389/fmicb.2017.02390.

(50) Small, P.; Blankenhorn, D.; Welty, D.; Zinser, E.; Slonczewski, J. L. *Acid and Base Resistance in Escherichia Coli and Shigella Flexneri: Role of RpoS and Growth PH*; 1994.

(51) Ohya, K.; Handa, Y.; Ogawa, M.; Suzuki, M.; Sasakawa, C. IpgB1 Is a Novel Shigella Effector Protein Involved in Bacterial Invasion of Host Cells: Its Activity to Promote Membrane Ruffling via Rac1 and Cdc42 Activation. *J. Biol. Chem.* **2005**, *280* (25), 24022–24034. https://doi.org/10.1074/jbc.M502509200.

(52) Hilbi, H.; Moss, J. E.; Hersh, D.; Chen, Y.; Arondel, J.; Banerjee, S.; Flavell, R. A.; Yuan, J.; Sansonetti, P. J.; Zychlinsky, A. Shigella-Induced Apoptosis Is Dependent on Caspase-1 Which Binds to IpaB. *J. Biol. Chem.* **1998**, *273* (49), 32895–32900. https://doi.org/10.1074/jbc.273.49.32895.

(53) Skoudy, A.; Mounier, J.; Aruffo, A.; Ohayon, H.; Gounon, P.; Sansonetti, P.; Van Nhieu, G. T. CD44 Binds to the Shigella IpaB Protein and Participates in Bacterial Invasion of Epithelial Cells. *Cell. Microbiol.* **2000**, *2* (1), 19–33. https://doi.org/10.1046/j.1462-5822.2000.00028.x.

(54) Watarai, M.; Funato, S.; Sasakawa, C. Interaction of Ipa Proteins of Shigella Flexneri with A5b1 Integrin Promotes Entry of the Bacteria into Mammalian Cells. *J. Exp. Med.* **1996**, *183* (March), 991–999.

(55) Ashida, H.; Mimuro, H.; Sasakawa, C. Shigella Manipulates Host Immune Responses by Delivering Effector Proteins with Specific Roles. *Front. Immunol.* **2015**, *6* (MAY), 1–12. https://doi.org/10.3389/fimmu.2015.00219.

(56) Fernandez-Prada, C. M.; Hoover, D. L.; Tall, B. D.; Hartman, A. B.; Kopelowitz, J.; Venkatesan, M. M. Shigella Flexneri IpaH7.8 Facilitates Escape of Virulent Bacteria from the Endocytic Vacuoles of Mouse and Human Macrophages. *Infect. Immun.* **2000**, *68* (6), 3608–3619. https://doi.org/10.1128/IAI.68.6.3608-3619.2000.

(57) Goldberg, M. B.; Theriot, J. A. Shigella Flexneri Surface Protein IcsA Is Sufficient to Direct Actin- Based Motility. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92* (14), 6572–6576. https://doi.org/10.1073/pnas.92.14.6572.

(58) Bernardini, M. L.; Mounier, J.; D'Hauteville, H.; Coquis-Rondon, M.; Sansonetti, P. J. Identification of IcsA, a Plasmid Locus of Shigella Flexneri That Governs Bacterial Intra-

and Intercellular Spread through Interaction with F-Actin. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86* (10), 3867–3871. https://doi.org/10.1073/pnas.86.10.3867.

(59) Krokowski, S.; Atwal, S.; Lobato-Márquez, D.; Chastanet, A.; Carballido-López, R.; Salje, J.; Mostowy, S. Shigella MreB Promotes Polar IcsA Positioning for Actin Tail Formation. *J. Cell Sci.* **2019**, *132* (9). https://doi.org/10.1242/jcs.226217.

(60) Maurelli, A. T.; Baudry, B.; Hélène D'hauteville, H.; Hale, T. L.; Sansonetti, P. J. Cloning of Plasmid DNA Sequences Involved in Invasion of HeLa Cells by Shigella Flexneri. *Infect. Immun.* **1985**, *49* (1), 164–171. https://doi.org/10.1128/iai.49.1.164-171.1985.

(61) Dorman, C. J.; McKenna, S.; Beloin, C. Regulation of Virulence Gene Expression in Shigella Flexneri, a Facultative Intracellular Pathogen. *Int. J. Med. Microbiol.* **2001**, *291* (2), 89–96. https://doi.org/10.1078/1438-4221-00105.

(62) Huang, J.; Brumell, J. H. Bacteria-Autophagy Interplay: A Battle for Survival. *Nat. Rev. Microbiol.* **2014**, *12* (2), 101–114. https://doi.org/10.1038/nrmicro3160.

(63) Adler, B.; Sasakawa, C.; Tobe, T.; Makino, S.; Komatsu, K.; Yoshikawa, M. A Dual Transcriptional Activation System for the 230 Kb Plasmid Genes Coding for Virulence-Associated Antigens of Shigella Flexneri. *Mol. Microbiol.* **1989**, *3* (5), 627–635. https://doi.org/10.1111/j.1365-2958.1989.tb00210.x.

(64) Porter, M. E.; Mitchell, P.; Roe, A. J.; Free, A.; Smith, D. G. E.; Gally, D. L. Direct and Indirect Transcriptional Activation of Virulence Genes by an AraC-like Protein, PerA from Enteropathogenic Escherichia Coli. *Mol. Microbiol.* **2004**, *54* (4), 1117–1133. https://doi.org/10.1111/j.1365-2958.2004.04333.x.

(65) Tobe, T.; Yoshikawa, M.; Mizuno, T.; Sasakawa, C. Transcriptional Control of the Invasion Regulatory Gene VirB of Shigella Flexneri: Activation by VirF and Repression by H-NS. *J. Bacteriol.* **1993**, *175* (19), 6142–6149.

(66) Prosseda, G.; Fradiani, P. A.; Di Lorenzo, M.; Falconi, M.; Micheli, G.; Casalino, M.; Nicoletti, M.; Colonna, B. A Role for H-NS in the Regulation of the VirF Gene of Shigella and Enteroinvasive Escherichia Coli. *Res. Microbiol.* **1998**, *149* (1), 15–25. https://doi.org/10.1016/S0923-2508(97)83619-4.

(67) Giangrossi, M.; Prosseda, G.; Tran, C. N.; Brandi, A.; Colonna, B.; Falconi, M. A Novel Antisense RNA Regulates at Transcriptional Level the Virulence Gene IcsA of Shigella Flexneri. *Nucleic Acids Res.* **2010**, *38* (10), 3362–3375. https://doi.org/10.1093/nar/gkq025.

(68) Tran, C. N.; Giangrossi, M.; Prosseda, G.; Brandi, A.; Di Martino, M. L.; Colonna, B.; Falconi, M. 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.* **2011**, *39* (18), 8122–8134. https://doi.org/10.1093/nar/gkr521.

(69) Gallegos, M. T.; Schleif, R.; Bairoch, A.; Hofmann, K.; Ramos, J. L. Arac/XylS Family of Transcriptional Regulators. *Microbiol. Mol. Biol. Rev.* **1997**, *61* (4), 393–410. https://doi.org/9409145.

(70) Egan, S. M. Growing Repertoire of AraC / XylS Activators. *J. Bacteriol.* **2002**, *184* (20), 5529–5532. https://doi.org/10.1128/JB.184.20.5529.

(71) Rhee, S.; Martin, R. G. R. G.; Rosner, J. L. J. L.; Davies, D. R. D. R. A Novel DNA-Binding Motif in MarA: The First Structure for an AraC Family Transcriptional

Activator. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (18), 10413–10418. https://doi.org/10.1073/pnas.95.18.10413.

(72) Kwon, H. J.; Bennik, M. H. J.; Demple, B.; Ellenberger, T. Crystal Structure of the Escherichia Coli Rob Transcription Factor in Complex with DNA. *Nat. Struct. Biol.* **2000**, *7* (5), 424–430. https://doi.org/10.1038/75213.

(73) Rhee, S.; Martin, R. G.; Rosner, J. L.; Davies, D. R. A Novel DNA-Binding Motif in MarA: The First Structure for an AraC Family Transcriptional Activator. *Proc.Natl.Acad.Sci.USA* **1998**, *95*, 10413–10418. https://doi.org/10.2210/PDB1BL0/PDB.

(74) Shi, J.; Wang, F.; Li, F.; Wang, L.; Xiong, Y.; Wen, A.; Jin, Y.; Jin, S.; Gao, F.; Feng, Z.; et al. Structural Basis of Transcription Activation by Rob, a Pleiotropic AraC/XylS Family Regulator. *Nucleic Acids Res.* **2022**, *50* (10), 5974–5987. https://doi.org/10.1093/nar/gkac433.

(75) Corbella, M.; Liao, Q.; Moreira, C.; Parracino, A.; Kasson, P. M.; Kamerlin, S. C. L. The N-Terminal Helix-Turn-Helix Motif of Transcription Factors MarA and Rob Drives DNA Recognition. *J. Phys. Chem. B* **2021**, *125* (25), 6791–6806. https://doi.org/10.1021/acs.jpcb.1c00771.

(76) Porter, M. E.; Dorman, C. J. 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.* **2002**, *184* (2), 531–539. https://doi.org/10.1128/JB.184.2.531-539.2002.

(77) Gillette, W. K.; Martin, R. G.; Rosner, J. L. Probing the Escherichia Coli Transcriptional Activator MarA Using Alanine-Scanning Mutagenesis: Residues Important for DNA

Binding and Activation. *J. Mol. Biol.* **2000**, *299* (5), 1245–1255.

https://doi.org/10.1006/jmbi.2000.3827.

Chapter 2 The VirF DNA Binding Domain

It has been well documented in the literature that the AraC family of transcriptional regulators are a very difficult family of proteins to work with *in vitro*.[1–5] These proteins are widespread within bacteria and are often involved in carbon metabolism, stress response, or virulence, making them interesting targets for antimicrobial inhibition.[6] To obtain any biochemical analysis of these proteins, researchers in the past have resorted to several tedious techniques to circumvent the insoluble nature of this protein family. The most commonly used technique is the denaturation and renaturation of the insoluble pellet containing the protein of interest through the use of buffers containing high concentrations of urea or guanidinium.[7–14] This process has been used to study AraC proteins such as MarA, Rv1395, RhaS, and SoxS, but refolding only yields a fraction of the overexpressed protein, often with variable levels of activity due to the nature of the refolding process.[12–15] Another approach is to create a fusion protein, attaching a maltose-binding protein (MBP) or N utilization substance A (NusA) tag to the AraC protein of interest.[8,16,17] There has been success in the literature using an MBP fusion tag to evaluate these proteins, but removal of the MPB tag often results in the protein reverting back to an insoluble form.[18,19] Additionally, the MPB tag is roughly 42 KDa, making it larger than the average size of the AraC proteins.[3,20] Taken together, the size of the tag and the inability to remove it without detrimental consequences to the solubility could potentially interfere with downstream function of the protein.[8,20]

To study VirF *in vitro*, our lab has spent much energy optimizing the purification process. Initial attempts to purify a MalE-VirF fusion protein (MBP-tagged) in various *E. coli* strains was

unsuccessful due to poor heterologous expression.[21] These expressions and purifications often lead to low yield and high impurity. To improve this method, our lab constructed an arabinose-inducible vector, pBAD202-MALvirF, which allowed for expression of MalE-VirF in *Shigella flexneri* BS103, an avirulent strain of the pathogen.[19] While this method was acceptable for the biochemical analysis, MalE-VirF is still not representative of native VirF. Although MalE-VirF has been shown to function in the *in vitro* analysis,[22] the dimerized active protein has approximately 85 KDa of excess mass at the N-terminal. To address this, we attempted to follow an established protocol from the literature that reported the successful purification of his-tagged VirF in *E. coli*.[23] Unfortunately, this method proved difficult to reproduce, as there was little to no expression of his-tagged VirF when following the published procedure.

To further study the VirF DNA binding domain, we wanted to purify a VirF construct that did not include the large MBP tag. Our first method was to revisit the previously mentioned published procedure that expresses full length VirF with a multi-histidine tag to try and optimize the yield. The second method was to express and purify a truncated form of the VirF C-terminal DNA binding domain. Here, we discuss those trials in detail.

Materials and Methods

Reagents and Plasmids

All reagents and standard buffer components were purchased from Thermo Fischer or Millipore Sigma unless otherwise specified in parentheses. DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT).

Table 2-1: Gene and Primer Sequences

Sequence Name	Sequence (5' → 3')
VirF gene	atgatggatatgggacataaaaacaaaatagatataaaggttcgcttgcataactatattattttat atgcaaaaaggtgttcaatgacggttagctcaggcaatgaaactttgactatcgatgaagggca aattgcttttatagagcgaaatatacaaataaacgtctccataaaaaaatctgatagcattaatcc atttgagattataagccttgacagaaatttattattaagcattattagaataatggaaccaatttattc atttcaacactcctattctgaggagaaaaggggttaaacaaaaaaatattcctcctctctgagg aggaggtttctatcgatttgttcaaatctataaaagagatgcctttcggcaaaagaaagatctata gtttagcttgccttttatcagctgtttctgatgaggaagctttatatacttcgatatcgatagcttcttctc ttagtttttctgatcagataaggaagattgttgaaaaaaacatcgagaagagatggcgtctttctg atatttcaaataacttgaatttatcagaaatagctgttagaaaacgattggagagtgaaaaattaa catttcaacaaatccttcttgatattcgcatgcatcatgcagcaaagcttttattgaatagtcaaagc tatattaatgatgtatcaagacttatcggaatatcaagcccatcttattttataaggaaatttaatga atattatggtataactccaaagaaatttttacttatatcataaaaaatttttaactcgagcaccaccac caccaccac
Primer Name	Sequence (5' → 3')
VirF AMP Forw	gagatccatatgatggacatgggac
VirF AMP Rev	tcggtcctcgaggaatttcttgtgatac
pET19b VirF opt forw	tatacatatgatggacatgggacataagaataaaatcgac
pET19b VirF opt rev	gtggtgggatccgaatttcttgtgatacaaataaaatttcttggg
VirF CDom Forw (144-263)	atgccatatgatggaggaagctttatatacttcg
VirF CDom Forw (164-263)	atgccatatgatgaggaagattgttgaaaaaaacatcg
VirF CDom Rev	ctcgagaaatttttatgatataagtaaaatttctttggag

His-tagged VirF Constructs – pET21b

A pET21b plasmid containing the *virF* gene with a C-terminal 6-histadine tag was constructed by a previous lab member. Upon sequencing the plasmid, a stop codon was recognized at the end of the *virF* gene, preventing the expression of the tag. To resolve this, a set of primers were ordered for mutagenesis (**Table II-1**). In a total volume of 50 µL, samples containing pET21bvirF-his (50 ng), forward and reverse PCR primers (0.5 µM each), dNTPs mix (2.5 mM each NTP), Phusion Polymerase Buffer (1x, New England Biolabs), and Phusion DNA Polymerase (1 U, New England Biolabs) were incubated according to the following temperature sequence: 1 cycle – 95°C for 10 minutes, 30 cycles – 95°C for 30 seconds, 52°C for 30 seconds, 72°C for 30 seconds, and 1 cycle – 72°C for 10 minutes before being incubated at 4°C overnight. The PCR product and the vector pET21b were separately subjected to a double

restriction enzyme digest by being treated with 10 U of *NdeI* and *XhoI* in Cutsmart® Buffer (New England Biolabs). The digested vector and PRC product were gel-purified from a 1% agarose gel and purified using a QIAquick Gel Extraction Kit (Qiagen). In a total volume of 20 µL, the digested *virF* PCR product and pET21b vector in a 5:1 ratio was combined with T4 ligase buffer and 1 U of T4 DNA Ligase (New England Biolabs) at 16°C overnight. The ligated product (pET21bvirF-6xHis) was transformed into 100 µL of chemically competent Top10 *E. coli* and plated on carbenicillin plates (50 µg/mL carbenicillin). Individual colonies were isolated and grown in 5 mL of 2xTY media (16 g bactotryptone, 10 g yeast extract, 5 g NaCl per liter of water supplemented with 100 µg/mL carbenicillin). The plasmid was purified using a Qiaprep Spin Miniprep Kit (Qiagen) and the removal of the stop codon was confirmed via DNA sequencing (University of Michigan DNA Sequencing Core Facilities).

His-tagged VirF Constructs – pET19b

In contrast to the pET21b vector, which contains a 6x-histidine tag downstream of *virF* gene, the pET19b vector harbors a 10x-histidine tag prior to the multiple cloning site, allowing for the incorporation of the tag on the N-terminal end of the gene. This would be beneficial when studying the VirF DNA binding domain by eliminating any possibility of the tag interfering with protein activity. To engineer this construct, *NdeI* and *BamHI* recognition sequences were inserted at the N-terminal and C-terminal of *virF* on pET21bvirF-6xHis, respectively, through PCR. In a total volume of 50 µL, samples containing pET21bvirF-6xHis (50 ng), forward and reverse PCR primers (0.5 µM each), dNTPs mix (2.5 mM each NTP), Taq Polymerase Buffer (1x, New England Biolabs), and Taq DNA Polymerase (1 U, New England Biolabs) were incubated according to the following temperature sequence: 1 cycle – 95°C for 10

minutes, 35 cycles – 95°C for 1 minute, 58°C for 1 minute, 72°C for 7 min, and 1 cycle – 72°C for 10 minutes before being incubated at 4°C overnight. To digest, 10 µL of Cutsmart Buffer and 1 U of both *NdeI* and *BamHI* were added to the reaction mixture before incubating at 37°C for 1.5 hours. To obtain the desired pET19b vector, pET19b-pp-CarD was obtained from Max Stefan of the Garcia Lab. The vector was digested by 10 U of *NdeI* and *BamHI* in Cutsmart Buffer at 37*C for 2 hours. Both the digested *virF* PCR product and pET19b vector were gel-purified from a 2% agarose gel and purified using a QIAquick Gel Extraction Kit (Qiagen). In a total volume of 20 µL, the digested *virF* PCR product and pET19b vector in a 3:1 ratio was combined with T4 ligase buffer and 1 U of T4 DNA Ligase (New England Biolabs) at 16°C overnight. The ligated product (pET19bvirFHis) was transformed into 100 µL of chemically competent Rosetta 2(DE3) pLysS *E. coli* and plated on carbenicillin plates (50 µg/mL carbenicillin). Individual colonies were isolated and grown in 5 mL of 2xTY media. The plasmid DNA was purified using a Qiaprep Spin Miniprep Kit (Qiagen) and the sequence was confirmed via DNA sequencing (University of Michigan DNA Sequencing Core Facilities).

pET21bvirF-6xHis Purification Trials

pET21bvirF-6xHis was transformed into chemically competent *E. coli* BL21(DE3) cells for expression and purification. An individual colony containing the plasmid was isolated and grown in 10 mL 2xTY media supplemented with carbenicillin (100 µg/mL) with agitation at 37°C overnight. A liter of fresh, sterile 2xTY was inoculated with the 10 mL starter culture and continued to be agitated at 37°C until the OD_{600} approximately read 0.5. Isopropyl β-d-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM and the culture was

allowed to grow overnight under the same conditions. The next day, the cells were harvested by centrifugation and the pellet was resuspended in 20 mL of lysis buffer (50 mM KH_2PO_4, 750 mM NaCl, 10 mM BME, 0.1% Tween-20, 10% glycerol, [pH 7.5]) supplemented with 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and a tablet of Roche cOmplete miniprotease inhibitor cocktail tablet (Roche). The solution was lysed via sonication and pelleted by centrifugation before the supernatant was collected and sterile filtered. The resulting solution was loaded onto a 1-mL HisTrap HP column (Cytiva) using an AKTA fast protein liquid chromatography (FPLC, GE Healthcare). The column was washed with 10 column volumes (CVs) of lysis buffer and the protein was eluted from the column with HisVirF elution buffer (50 mM KH_2PO_4, 500 mM NaCl, 500 mM imidazole, 10 mM BME, 0.1% Tween-20, 20% glycerol, [pH 7.5]). Fractions were collected and analyzed on a 10% polyacrylamide gel and fraction 2 was chosen for further purification. The solution was transferred to a cation exchange (CIEX) binding buffer (10 mM HEPES, 5 mM BME, 0.1% Tween-20, 20% glycerol, [pH 7.5]) and was loaded onto a Source 15Q (Cytiva) CIEX column. The protein was eluted with increasing concentration of CIEX elution buffer (bind buffer with 1 M NaCl) and the fractions were visualized by SDS PAGE. No bands were detected.

A colony of chemically competent *E. coli* BL21(DE3) was isolated and grown in 10 mL 2xTY media supplemented with carbenicillin (100 µg/mL) with agitation at 37°C overnight. A liter of fresh, sterile 2xTY was inoculated with the 10 mL starter culture and continued to be agitated at 37°C until the OD_{600} approximately read 0.5. IPTG was added to a final concentration of 1 mM and the culture was allowed to grow overnight under the same conditions. The next day, the cells were harvested by centrifugation and the pellet was resuspended in 20 mL of lysis buffer (as previously described, adding 10% glycerol, 2% D-mannitol) supplemented

with 0.1 mM PMSF and a tablet of Roche cOmplete miniprotease inhibitor cocktail tablet (Roche). The solution was lysed via sonication and pelleted by centrifugation. At this stage, the pellet was collected instead of the supernatant and was resuspended in denaturing buffer (20 mM Tris HCl, 500 mM NaCl, 6 M guanidinium chloride, [pH 7.5]) before centrifugation. The cell lysate was applied to a 1-mL HisTrap HP column (Cytiva) using an AKTA FPLC column (GE Healthcare). The protein was eluted with increasing concentrations of denaturing elution buffer (denaturing buffer with 500 mM imidazole) and each stage of the purification was visualized with SDS PAGE. The protein appeared to be trapped in the pellet, leading us to subject the pellet to denaturing buffer overnight. The process was repeated, applying the solution to another 1-mL HisTrap HP column, but after SDS PAGE no protein was detected in the eluted fractions.

pET19bvirFHis Purification Trial

To purify His-tagged VirF, pET19bvirFHis was transformed into chemically competent *E. coli* BL21(DE3) cells for expression and purification. Similar to as previously described, an individual colony containing the plasmid was isolated and grown in 10 mL 2xTY media supplemented with carbenicillin (100 µg/mL) with agitation at 37°C overnight. A liter of fresh, sterile 2xTY was inoculated with the 10 mL starter culture and continued to be agitated at 37°C until the OD_{600} approximately read 0.5 and the culture was induced with 1 mM IPTG overnight at 16°C. The culture was pelleted by centrifugation and resuspended in lysis buffer (20 mM Tris HCl, 300 mM NaCl, 5% glycerol, 0.1% Tween-20, 10 mM BME, [pH 7.5]). The solution was lysed via sonication and the soluble lysate was separated by centrifugation. The lysate was applied to a HisTrap HP 1mL column on the AKTA FPLC and the protein was eluted with

increasing concentrations of elution buffer (lysis buffer with 500 mM imidazole and 20% glycerol) on SDS PAGE. Samples thought to contain protein were pooled, flash frozen, and stored in a -80°C freezer.

Electrophoretic Mobility Shift Assay

To evaluate His-tagged VirF's capability to bind DNA, an electrophoretic mobility shift assay (EMSA) was performed. The EMSA protocol was adapted from Emanuele et al 2015.[19] Briefly, reactions consisting of 6 µL *pvirB* EMSA DNA probe (0.25 µM), 6 µL of either His-tagged VirF (1.0 µM) or native gel loading buffer (300 mM Tris HCl, 50% glycerol, 0.05% bromophenol blue, [pH 7.0]), 1 µL of salmon sperm DNA (0.7 mg/mL, Invitrogen), and 0.5 µL BSA (0.07 mg/mL) were incubated at 37°C in a water bath for 15 minutes. A 6% native polyacrylamide gel (29:1 acrylamide to bis-acrylamide ratio) was made with 0.25X TBE buffer (22 mM Tris Base, 22 mM boric acid, 0.5 mM EDTA, [pH 9.5]) for the EMSA. The gel was electrophoresed for 1 hour at 150 V in 0.25X TBE buffer at 4°C before samples were loaded. After the reaction solutions (12 µL) were loaded onto the gel, the gel was electrophoresed for an additional hour at 150 V and 4°C. Gel visualization was performed using a Molecular Dynamics Typhoon 9200 molecular imager by excitation (Ex) at 607 nm and reading the 710-nm emission (Em).

VirF DNA Binding Domain Constructs

To truncate VirF, primers were designed that would clone *virF* starting at either E144 (144-262) or R164 (164-262) from pET21bvirF-6xHis and into a TOPO vector (**Table II-1**). In

a total volume of 50 µL, two separate PCR reactions containing forward and reverse primers (0.5 µM, either E144 or R164 forward primer), pET21bvirF-6xHis (50 ng) dNTPs mix (2.5 mM each NTP) Taq Polymerase Buffer (1x, New England Biolabs), and Taq DNA Polymerase (1 U, New England Biolabs) were incubated according to the following temperature sequence: 1 cycle – 95°C for 10 minutes, 35 cycles – 95°C for 1 minute, 57°C for 1 minute, 72°C for 7 min, and 1 cycle – 72°C for 10 minutes before being incubated at 4°C overnight. The next day, the PCR product was cloned into a TOPO Vector according to manufacturer's protocol and transformed into TOP10 *E. coli*. TOPO Vectors containing each truncated *virF* gene were isolated and digested in 10 µL of Cutsmart Buffer and 1 U of both *NdeI* and *XhoI* were added to the reaction mixture before incubating at 37°C for 1.5 hours. The pET19b vector was simultaneously digested by 1 U of both *NdeI* and *XhoI* in Cutsmart Buffer for 2 hours. Both digested *virF* products and pET19b vector were gel-purified from a 2% agarose gel and purified using a QIAquick Gel Extraction Kit (Qiagen). In a total volume of 20 µL, the digested *virF* product and pET19b vector in a 3:1 ratio was combined with T4 ligase buffer and 1 U of T4 DNA Ligase (New England Biolabs) at 16°C overnight. The ligated products (pET19bVirF144-262 and pET19bVirF164-262) were transformed into 100 µL of chemically competent BL21(DE3) *E. coli* and plated on carbenicillin plates (50 µg/mL carbenicillin). Individual colonies were isolated and grown in 5 mL of 2xTY media. The plasmid DNA was purified using a Qiaprep Spin Miniprep Kit (Qiagen) and the sequence was confirmed via DNA sequencing (University of Michigan DNA Sequencing Core Facilities).

Expression Test

Comprehensive expression tests were performed to determine the level of expression that is achieved with both pET19bVirF144-262 and pET19bVirF164-262 in *E. coli*. First, 10 mL cultures of 2xTY containing *E. coli* housing one of the two expression plasmids were grown overnight. These were used to inoculate cultures containing fresh 2xTY and were allowed to continue growing at 37°C until OD_{600} reached 0.5. One set of cultures were then induced with 0.5 mM IPTG and grew for various induction times (2, 3, 4, or 16 hours) and the other set was induced with varying amounts of IPTG (0.1 mM, 0.25 mM, 0.5 mM, or 1 mM) for 3 hours. At each time point, 1 mL of the culture was collected, pelleted by centrifugation, and resuspended in 100 µL of lysis buffer (20 mM Tris HCl, 250 mM NaCl, 10 mM BME, 0.1% Triton X-100, 5 mM imidazole, 30% glycerol, [pH 7.9]). Expression of each sample was visualized by SDS PAGE.

Purification Trials of the VirF C-Terminal DNA Binding Domain Constructs

Both pET19bVirF144-262 and pET19bVirF164-262 were expressed and purified under the same conditions side by side. An individual colony of BL21(DE3) *E. coli* containing one of the two plasmids was isolated and grown in 10 mL 2xTY media supplemented with carbenicillin (100 µg/mL) with agitation at 37°C overnight. A liter of fresh 2xTY was sterilized for each plasmid and was inoculated with the appropriate 10 mL starter culture. The flask continued to be agitated at 37°C until the OD_{600} approximately read 0.5. IPTG was added to a final concentration of 1 mM and the culture was allowed to grow overnight under the same conditions. The next day, the cells were harvested by centrifugation and the pellet was resuspended in 20 mL of lysis buffer (same as expression test) supplemented with 0.1 mM PMSF and a tablet of Roche cOmplete miniprotease inhibitor cocktail tablet (Roche). The solution was lysed via sonication

and pelleted by centrifugation. The lysates were collected and sterile filtered before they were applied to a 1-mL HisTrap HP column (Cytiva) using an AKTA FPLC. The column was washed with 10 C.V. of lysis buffer before applying increasing concentrations of elution buffer (lysis buffer with 500 mM imidazole). The fractions that were thought to contain the VirF DNA binding domains (DBD) according to the chromatograms were collected and analyzed on a 10% polyacrylamide gel. Unfortunately, neither of the two proteins were identified in the gel.

Next, we tried to purify the VirF DBD using a HiTrap SP FF cation exchange column (Cytiva). Obtaining the cell pellet was achieved as described above. Once pelleted, the cells are resuspended in CIEX bind buffer and lysed via sonication. The solution was centrifuged, and the lysate was filtered and applied the HiTrap SP FF CIEX using AKTA FPLC. The protein was eluted with increasing concentration of CIEX elution buffer. The fractions that were thought to contain the VirF DBDs according to the chromatograms were collected and analyzed on a 14% polyacrylamide gel. VirF DBD 144-262 was observed in the fractions corresponding to the first elution peak on the chromatogram, but if VirF DBD 164-262 was present, the samples were too impure to see. Activity of VirF DBD 144-262 was evaluated by EMSA as previously described. Going forward, only pET19bVirF144-262 was use in the expression trials unless otherwise stated. To obtain pure and active VirF DBD, we tried to co express the DBD in the presence of the protein chaperones GroEL/ES. The reason was that these chaperones may help fold the VirF DBD into the proper tertiary structure without us having to denature protein. First, pET19bVirF144-262 was transformed into chemically competent *E. coli* BL21(DE3) cells housing the expression plasmid pGro7 with GroEL/ES donated by Dr. Emily Scott's lab. An individual colony containing the plasmids was isolated and grown in 10 mL 2xTY media supplemented with carbenicillin (100 µg/mL) with agitation at 37°C overnight. 500 mL of fresh

2xTY was inoculated with the starter culture and continued to be agitated at 37°C. When the OD_{600} read 0.3, 5 mL of 20% arabinose was added and the culture continued to grow under the same conditions. When the OD_{600} read 0.6, the culture was induced with 0.4 mM IPTG and grew overnight at 16°C. The next day, the cells were harvested by centrifugation and the pellet was resuspended in 25 mL of lysis buffer (same as *pET19bvirFHis* purification) supplemented with 0.1 mM PMSF and a tablet of Roche cOmplete miniprotease inhibitor cocktail tablet (Roche). The solution was lysed via sonication and pelleted by centrifugation before the supernatant was collected and sterile filtered. The resulting solution was loaded onto a 5-mL HisTrap HP column (Cytiva) using an AKTA FPLC. The column was washed with 10 column volumes (CVs) of lysis buffer and the protein was eluted from the column with increasing concentration of elution buffer. Fractions were collected and analyzed on a 14% polyacrylamide gel. Almost all the over expressed VirF DBD was in the pellet.

To express and purify the VirF DBD from the insoluble pellet, we modified a published protocol used to purify the *E. coli* AraC transcriptional regulator, MarA.[12] First, a culture of *E. coli* BL21(DE3) containing pET19bVirF144-262 was grown overnight at 37°C and was used to inoculate a fresh 1 L culture of 2xTY. The culture was grown at 37°C with agitation until the OD_{600} was approximately 0.8. The culture was induced with IPTG at a final concentration of 0.4 mM and vigorously rocked at 37°C for 3 more hours before the cells were pelleted. The pellet was then rinsed with 25 mL of lysis buffer (50 mM Tris-HCl, 1 mM EDTA, 1 M NaCl [pH 7.5]) before freezing overnight. The next morning, the cells were resuspended in 25 mL of lysis buffer and lysed via sonication. The solution was pelleted by ultracentrifugation (120,000 x g, 4°C, 30 min), the supernatant was discarded, and the pellet was washed with 30 mL of denture buffer 1 (50 mM Tris-HCl, 4 M urea [pH 8.5]) before the ultracentrifugation was repeated. The

supernatant was discarded again, and the pellet was resuspended in 25 mL of denature buffer 2 (50 mM Tris-HCl, 6 M guanidinium chloride [pH 8.5]). The mixture was subjected to ultracentrifugation a third time, and the supernatant was collected and diluted to 40 mL with resuspension buffer (50 mM Tris-HCl [pH 8.5]). The solution was loaded onto a HisTrap 5 mL column using an AKTA FPLC and washed with 3 C.V. of water. The protein was eluted by applying a linear gradient of 0-100% wash buffer (50 mM Tris-HCl, 500 M NaCl, [pH 8.5]) to elution buffer (wash buffer with 1 M imidazole]). The fractions that eluted around 0.2 M imidazole were analyzed on SDS PAGE and the fractions containing a band expected to be the VirF DBD were pooled and dialyzed overnight (50 mM HEPES, 1 M NaCl, 1 mM DTT, 5 mM EDTA, 0.1 mM Triton X-100, 20% glycerol, [pH 8.0]). The solution was concentrated, aliquoted, and flash frozen in liquid nitrogen. The DNA binding activity of the protein was evaluated by EMSA as previously described.

Suspicious that the purified VirF DBD 144-262 was inactive, we attempted to refold the purified protein on the nickel column. Briefly, a 1 mL purified protein stock of VirF DBD 144-262 was denatured in 10 mL of denature buffer 2 and rocked at 4°C overnight. The following day, the solution was applied to a HisTrap 5 mL column with ATKA FPLC. The column was washed with wash buffer supplemented with 6 M guanidinium chloride, slowly decreasing the concentration of guanidinium chloride to 0 M of 20 C.V. The protein was eluted by applying a linear gradient of 0-100% wash buffer to elution buffer. The collected fractions were visualized with SDS PAGE.

Results

pET21bVirF-6xHis

Initial attempts to express and purify VirF from a pET21b vector were unsuccessful due to a stop codon that was mistakenly placed at the end of the gene. Using the VirF AMP Forw/Rev primers from **Table II-1**, the stop codon was removed after troubleshooting the PCR conditions. We first tried to purify VirF-6xHis with FPLC on a nickel column modifying a published procedure.[23] At first, we thought the results looked promising, identifying a band in fraction 1 and 2 that may be our 31 Kd protein **(Figure 2-1)**. Unfortunately, after attempting to clean up the fraction on a subsequent cation exchange column (CIEX), no protein appeared to elute from the column. Following this, we repeated the purification under denaturing conditions. Guanidinium chloride was added to the solution when we resuspended the pellet and remained present through eluting from the column. After visualizing the purification steps with SDS PAGE, it was apparent that VirF-6xHis was still trapped in the pellet and what we saw in the previous purification was misidentified **(Figure 2-2)**.

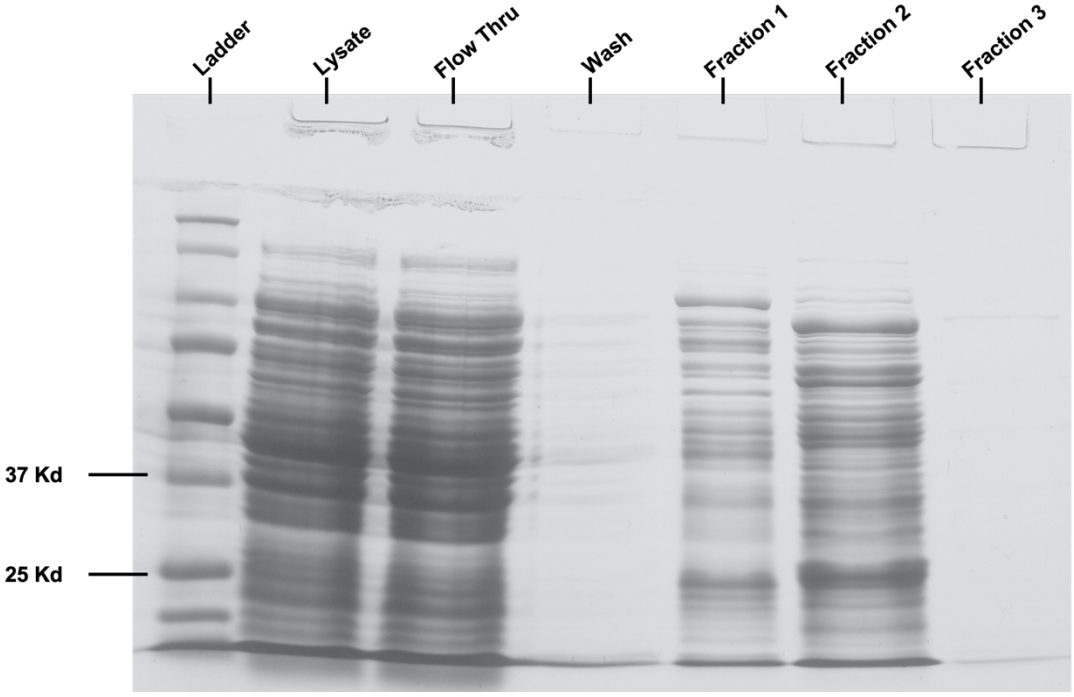


Figure 2-1: ***Purification Gel of VirF-6xHis Post Ni-Affinity Chromatography***

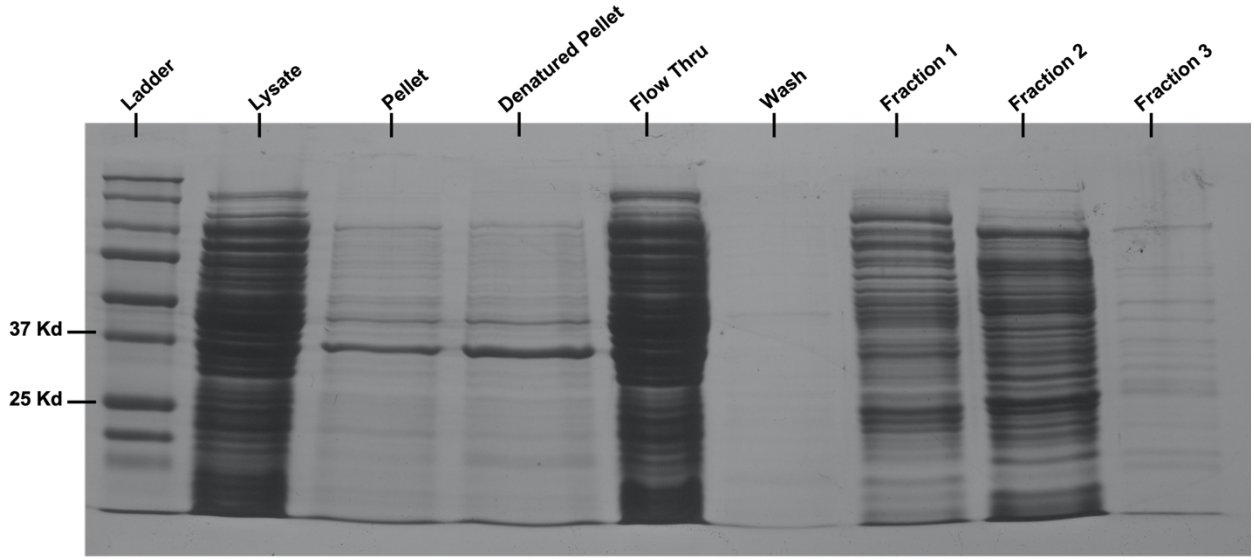


Figure 2-2: ***SDS PAGE of VirF-6xHis Denaturing Purification: The protein appears to be trapped in the pellet****.*

pET19bVirFHis

Going forward in our attempts to purify VirF, we decided to change expression vectors from pET21b to pET19b. This would allow us to have a 10x histidine tag on the N-terminal end of the protein instead of at the C-terminal. It is foreseeable that removal of the tag may be detrimental to protein solubility, and this would keep it from interfering with the activity of the C-terminal DNA binding domain. Additionally, 10x histidine tags may be more compatible with downstream assays such as SPR.[24] This was achieved using pET19b VirF Opt Forw/Rev primers after several round of optimization and DNA sequencing. After much troubleshooting, the purification was finally able to isolate low yields of 10xHis-VirF **(Figure 2-3)**. The activity of the protein was evaluated by an electrophoretic mobility shift assay (EMSA). The native gel showed no shift that would indicate any binding to the DNA by 10xHis VirF. The experiment was repeated several times but binding by 10xHis VirF was never observed.

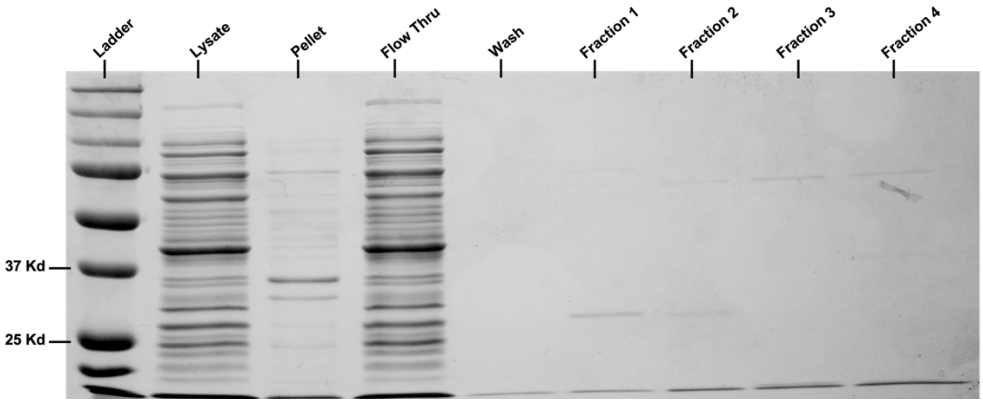


Figure 2-3: ***SDS PAGE of 10xHis VirF Purification***

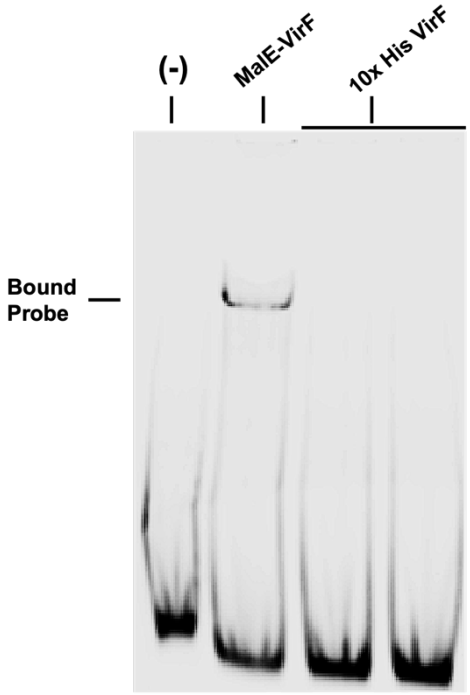


Figure 2-4: ***EMSA Evaluating 10xHis VirF Binding to a virB Probe***

VirF DNA Binding Domain

We truncated *virF* to obtain two constructs of the VirF DNA binding domain (DBD) of different lengths referred to as VirF DBD 144-262 or VirF DBD 164-262 **(Table 2-2)**. This was

achieved by subcloning the truncated PCR products into a TOPO vector using the primers described in **Table 2-1**. The genes were ligated into a pET19b vector for expression and purification and were ready after correcting mutations on pET19bVirF164-262 that were a byproduct of the cloning. The first few attempts to express and purify these two truncated proteins looked to fail due to low expression, despite our expression tests prior to the purification. Even though the chromatograms showed peaks where protein should be eluting, the purification gel in **Figure 2-5** failed to show any bands near the predicted protein masses (VirF DBD 144-262 = 16 Kd; VirF DBD 164-262 = 13 Kd).

Table 2-2: ***VirF amino acid sequence depicting the starting points for each of the truncated DNA binding domains.***

Sequence Name	Amino Acid Sequence
VirF 144-262 164-262	MMDMGHKNKIDIKVRLHNYIILYAKRCSMTVSSGNETLTIDEGQIAFIER NIQINVSIKKSDSINPFEIISLDRNLLLSIIRIMEPIYSFQHSYSEEKRGLNK KIFLLSEEEVSIDLFKSIKEMPFGKRKIYSLACLLSAVSDEEALYTSISIAS SLSFSDQIRKIVEKNIEKRWRLSDISNNLNLSEIAVRKRLESEKLTFQQIL LDIRMHHAAKLLLNSQSYINDVSRLIGISSPSYFIRKFNEYYGITPKKFYL YHKKF

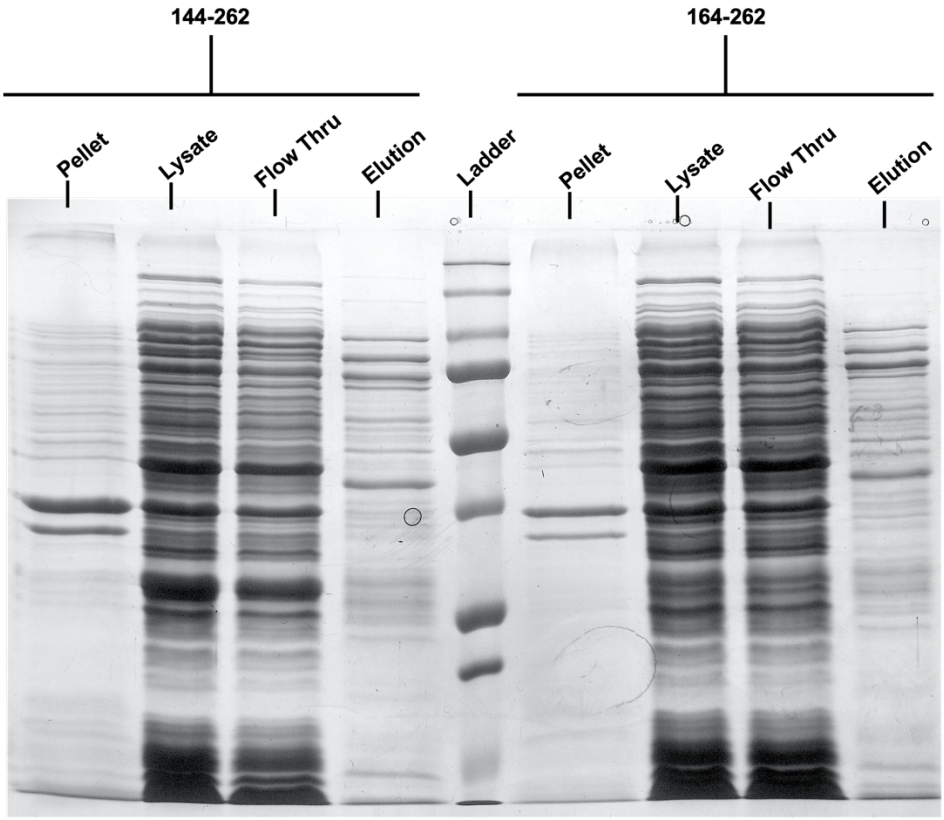


Figure 2-5: ***First Purification Gel of the VirF DBDs***

When using a CIEX column, we had better luck isolating VirF DBD 144-262. Both constructs were expressed and purified on the FPLC, but VirF DBD 164-262 was much more impure after elution **(Figure 2-6)**. The activity of VirF DBD 144-262 was evaluated by an EMSA. Purified MalE-VirF was used as a positive control and showed binding to the *virB* probe, but no shift was observed in either of the two lanes corresponding to the purified DBD **(Figure 2-7)**. This led us to focus on cleaning the purification of VirF DBD 144-262.

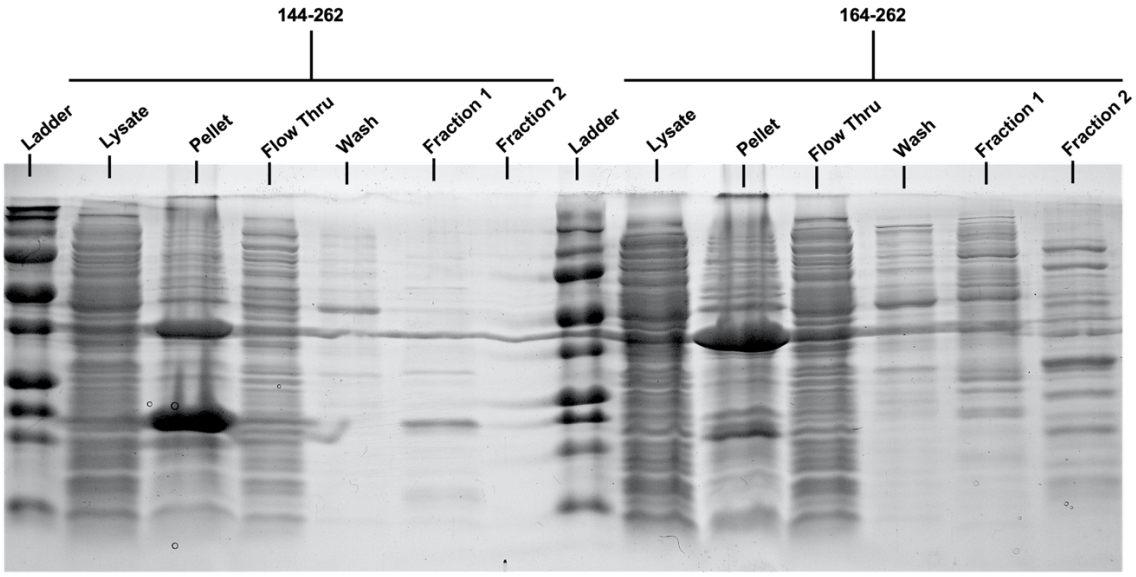


Figure 2-6: ***CIEX Purification Gel of DBDs.*** *Fraction 1 of VirF DBD 144-262 looks to contain our protein.*

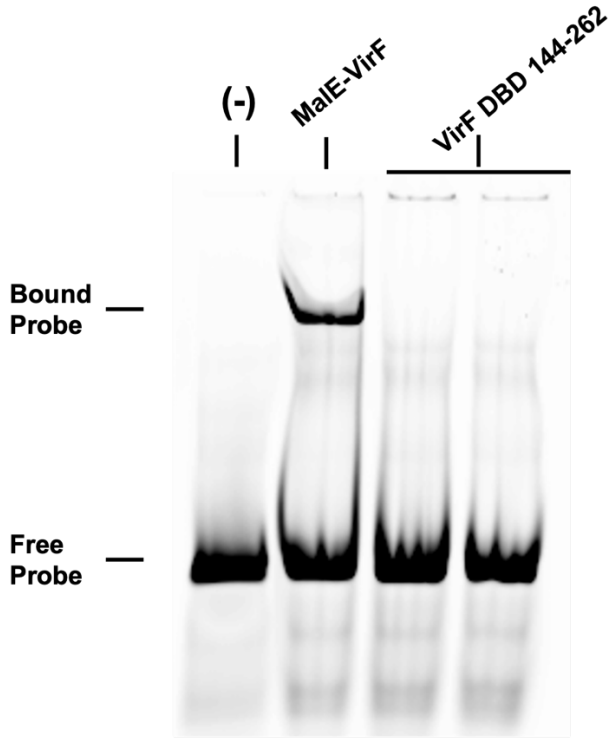


*Figure 2-7:****EMSA Gel to Test VirF DBD 144-262 Activity****. No binding observed in the duplicate reactions.*

We co-expressed the chaperones GroEL/ES with VirF DBD 144-262 with the expectation that the chaperones will help facilitate the proper folding to the tertiary structure.

Unfortunately, the eluted fraction showed no protein when visualized by SDS PAGE. In fact, almost all the expressed VirF DBD 144-262 appeared to be trapped in the insoluble pellet **(Figure 2-8)**. To try and purify from the insoluble pellet, we modified a protocol used by Jair et al. to isolate active MarA.[12] This protocol resulted in a significantly higher yield and a relatively pure sample when visualized by SDS PAGE **(Figure 2-9)**. Again, the DNA binding activity was evaluated by EMSA using MalE-VirF as a positive control but no shift indicating binding was observed for VirF DBD 144-262 **(Figure 2-10)**. We tried to refold the protein while bound to the nickel column and again were able to isolate a decent yield in a clean fraction. Unfortunately, when we tested the activity, the protein showed no binding activity.

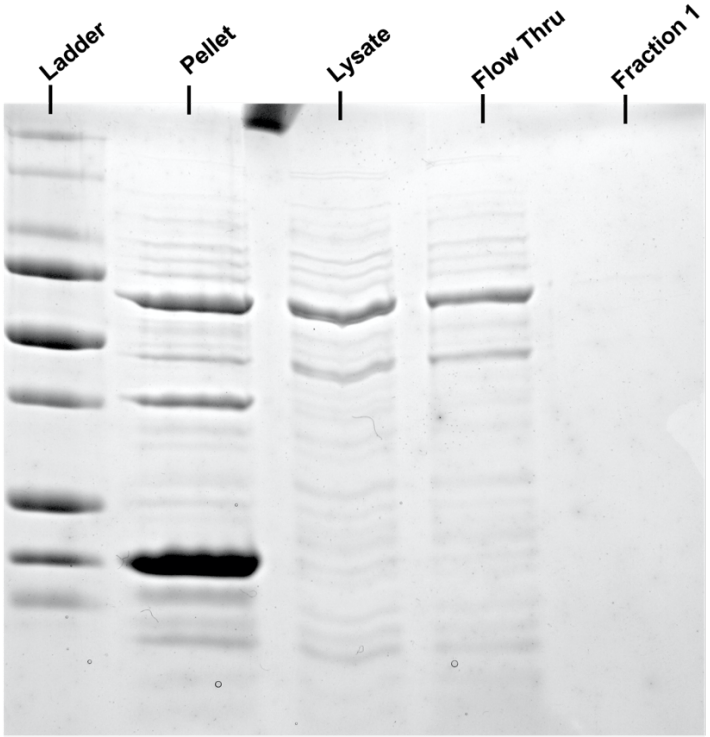


Figure 2-8: ***Purification Gel of VirF DBD 144-262 Using GroEL/ES.*** *Protein appears to be trapped in the pellet*

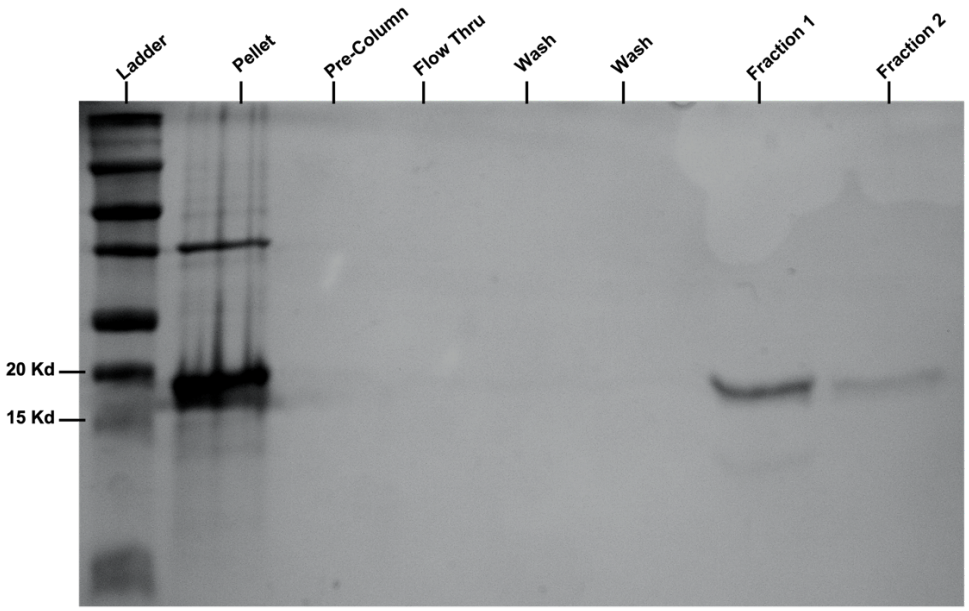


Figure 2-9: VirF DBD 144-262 Purification from the Insoluble Pellet

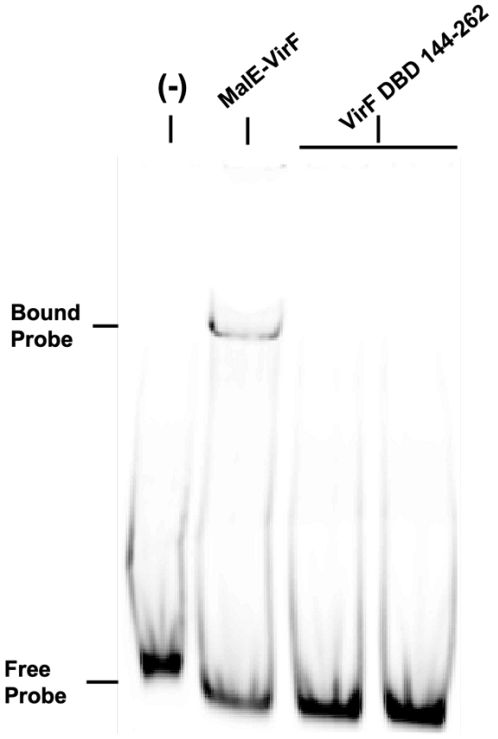


Figure 2-10: ***EMSA to Test DNA Binding Activity of VirF DBD 144-262***

Discussion

In *Shigella*, VirF is a crucial component to pathogenicity. It's ability to control the expression of key virulence factors makes VirF a compelling protein for investigation.[25,26] The protein is made up of two distinct domains: a N-terminal dimerization domain and a C-terminal DNA binding domain. The helix-turn-helix DNA binding domain is a conserved feature across all AraC family members, suggesting significance of this feature across a wide variety of bacterial species.[3] Similar to other AraC family members, the insolubility of VirF when expressed in *E. coli* makes analysis of the structure and function *in vitro* challenging. While the use of an MBP tag on VirF assists with solubility, the size of the tag leaves questions about the biological relevance of the protein.

In 2011, Tran et al. published a manuscript that detailed experiments using a his-tagged wild-type VirF.[23] This was the first reported purification of VirF that did not use a MBP tag, allowing for the analysis of the protein without concern of the excess mass hindering the activity. When we tried to repeat the procedure with pET21bVirF-6xHis, we continuously found our protein would be trapped in the insoluble pellet. When using different buffers throughout the process failed to improve protein solubility, we thought that the denaturation of the protein in the pellet would help remove VirF-6xHis from the inclusion bodies. Unfortunately, seen in **Figure 2-2**, most of the expressed protein is still in the pellet, despite it being resuspended in high concentration of guanidinium chloride.

We decided that isolating a construct of VirF with an N-terminal tag might be more beneficial than having a C-terminal tag. Our lab already had several other expression constructs that used pET19b as a vector. While there were many issues attempting to clone the gene from one vector to the other, we eventually were able to verify that our expression plasmid

pET19bVirFHis was correct. Just as before, the purification was repeated several times to develop a procedure that yielded VirF. This trial-and-error approach was necessary due to many purifications failing to show any protein eluting from the column. In the end, the wash and elution buffer components listed in the methods for this purification were successful in getting the protein to elute off the column where expected. Despite finally being able to isolate a relatively decent yield, 10xHis VirF showed no activity in the EMSA. The same *virB* probe was used for both the MalE-VirF control and the 10xHis VirF test but only the MalE-VirF demonstrated the ability to bind to DNA.

Since other members of the AraC family of transcription activators show better solubility than VirF, we thought that isolating a truncated form of the VirF DBD would have more success *in vivo*.[3,27] Studying a predicted model of full length VirF, we chose two places to truncate the linker between the DNA binding domain and the dimerization domain **(Table 2-2)**. At first, we chose to put our energy into optimizing the expression and purification of both truncated forms (VirF DBD 144-262 and VirF DBD 164-262). We reasoned that it would be beneficial to pursue clean isolates of both versions to compare the activity of the two, and we tried standard affinity chromatography and ion exchange chromatography to achieve this. We would consistently see low expression but were able to isolate VirF 144-262 following CIEX **(Figure 2-6)**. After the purification of VirF 164-262, we saw a band that may correspond to the small 13 Kd protein, but the eluted fraction was very impure compared to VirF DBD 144-262. When the impure fraction was reapplied to the column, the impurities were not lessened. Once we tested the activity of the isolated VirF 144-262 and saw that it was in capable of binding to DNA **(Figure 2-7)**, we decided to focus on improving the purification of this truncated form over trying to clean up the purification of VirF 164-262.

We reasoned that the protein is being misfolded at some stage of the purification, leading us to try expressing the DBD in the presences of protein chaperones GroEL/ES. Although the chromatogram showed a peak during the elution, it was small and the SDS PAGE showed all the expressed protein was still trapped in the cell pellet. Several AraC proteins have been isolated from a cell pellet in a denaturing purification, but this often hinders protein activity and yield. Despite this, we chose to modify a protocol used by Jair et al. to purify active MarA.[12–15] We were surprised to see that this method resulted in a relatively high yield of VirF DBD when compared to the yield that was achieved when not denaturing the protein. However, as predicted, achieving pure and active protein through this method still proved difficult. While each protein purification we tried through this method resulted in decent yield, none of the trials produced protein that was active in the DNA binding assay **(Figure 2-10)**. Going forward, evaluating the expression of different lengths of truncated versions of VirF may prove to be more successful than what we were able to accomplish. VirF remains a difficult protein to work with *in vivo* and any analysis will have to be achieved with MalE-VirF for the time being.

References

(1) Schleif, R. AraC Protein: A Love-Hate Relationship. *BioEssays* **2003**, *25* (3), 274–282. https://doi.org/10.1002/bies.10237.

(2) Porter, M. E.; Dorman, C. J. 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.* **2002**, *184* (2), 531–539. https://doi.org/10.1128/JB.184.2.531-539.2002.

(3) Gallegos, M. T.; Schleif, R.; Bairoch, A.; Hofmann, K.; Ramos, J. L. Arac/XylS Family of Transcriptional Regulators. *Microbiol. Mol. Biol. Rev.* **1997**, *61* (4), 393–410. https://doi.org/9409145.

(4) Ruíz, R.; Marqués, S.; Ramos, J. L. Leucines 193 and 194 at the N-Terminal Domain of the XylS Protein, the Positive Transcriptional Regulator of the TOL Meta-Cleavage Pathway, Are Involved in Dimerization. *J. Bacteriol.* **2003**, *185* (10), 3036–3041. https://doi.org/10.1128/JB.185.10.3036-3041.2003.

(5) Timmes, A.; Rodgers, M.; Schleif, R. Biochemical and Physiological Properties of the DNA Binding Domain of AraC Protein. *J. Mol. Biol.* **2004**, *340* (4), 731–738. https://doi.org/10.1016/j.jmb.2004.05.018.

(6) Roncarati, D.; Scarlato, V.; Vannini, A. Targeting of Regulators as a Promising Approach in the Search for Novel Antimicrobial Agents. *Microorganisms* **2022**, *10* (1), 1–16. https://doi.org/10.3390/microorganisms10010185.

(7) Egan, S. M.; Schleif, R. F. DNA-Dependent Renaturation of an Insoluble DNA Binding Protein: Identification of the RhaS Binding Site at RhaBAD. *Journal of Molecular Biology*. 1994, pp 821–829. https://doi.org/10.1006/jmbi.1994.1684.

(8) Structural Genomics Consortium. Protein Production and Purification. *Nat. Methods* **2008**, *5* (2), 135–146. https://doi.org/10.1038/nmeth.f.202.Protein.

(9) Vincentelli, R.; Canaan, S.; Campanacci, V.; Valencia, C.; Maurin, D.; Frassinetti, F.; Scappucini-Calvo, L.; Bourne, Y.; Cambillau, C.; Bignon, C. High-Throughput Automated Refolding Screening of Inclusion Bodies. *Protein Sci.* **2009**, *13* (10), 2782–2792. https://doi.org/10.1110/ps.04806004.

(10) Willis, M. S.; Hogan, J. K.; Prabhakar, P.; Liu, X.; Tsai, K.; Wei, Y.; Fox, T. Investigation of Protein Refolding Using a Fractional Factorial Screen: A Study of Reagent Effects and Interactions. *Protein Sci.* **2005**, *14* (7), 1818–1826. https://doi.org/10.1110/ps.051433205.

(11) Gillette, W. K.; Martin, R. G.; Rosner, J. L. Probing the Escherichia Coli Transcriptional Activator MarA Using Alanine-Scanning Mutagenesis: Residues Important for DNA Binding and Activation. *J. Mol. Biol.* **2000**, *299* (5), 1245–1255. https://doi.org/10.1006/jmbi.2000.3827.

(12) Jair, K. W.; Martin, R. G.; Rosner, J. L.; Fujita, N.; Ishihama, A.; Wolf, R. E. Purification and Regulatory Properties of MarA Protein, a Transcriptional Activator of Escherichia Coli Multiple Antibiotic and Superoxide Resistance Promoters. *J. Bacteriol.* **1995**, *177* (24), 7100–7104. https://doi.org/10.1128/jb.177.24.7100-7104.1995.

(13) Recchi, C.; Sclavi, B.; Rauzier, J.; Gicquel, B.; Reyrat, J. M. Mycobacterium Tuberculosis Rv1395 Is a Class III Transcriptional Regulator of the AraC Family Involved in Cytochrome P450 Regulation. *J. Biol. Chem.* **2003**, *278* (36), 33763–33773. https://doi.org/10.1074/jbc.M305963200.

(14) Wickstrum, J. R.; Skredenske, J. M.; Kolin, A.; Jin, D. J.; Fang, J.; Egan, S. M. Transcription Activation by the DNA-Binding Domain of the AraC Family Protein RhaS

in the Absence of Its Effector-Binding Domain. *J. Bacteriol.* **2007**, *189* (14), 4984–4993. https://doi.org/10.1128/JB.00530-07.

(15) Li, Z.; Demple, B. SoxS, an Activator of Superoxide Stress Genes in Escherichia Coli. Purification and Interaction with DNA. *J. Biol. Chem.* **1994**, *269* (28), 18371–18377. https://doi.org/10.1016/s0021-9258(17)32317-7.

(16) Nallamsetty, S.; Waugh, D. S. A Generic Protocol for the Expression and Purification of Recombinant Proteins in Escherichia Coli Using a Combinatorial His6-Maltose Binding Protein Fusion Tag. *Nat. Protoc.* **2007**, *2* (2), 383–391. https://doi.org/10.1038/nprot.2007.50.

(17) Nallamsetty, S.; Waugh, D. S. Solubility-Enhancing Proteins MBP and NusA Play a Passive Role in the Folding of Their Fusion Partners. *Protein Expr. Purif.* **2006**, *45* (1), 175–182. https://doi.org/10.1016/j.pep.2005.06.012.

(18) Shin, S.; Castanie-Cornet, M. P.; Foster, J. W.; Crawford, J. A.; Brinkley, C.; Kaper, J. B. An Activator of Glutamate Decarboxylase Genes Regulates the Expression of Enteropathogenic Escherichia Coli Virulence Genes through Control of the Plasmid-Encoded Regulator, Per. *Mol. Microbiol.* **2001**, *41* (5), 1133–1150. https://doi.org/10.1046/j.1365-2958.2001.02570.x.

(19) Emanuele, A. A.; Garcia, G. A. Mechanism of Action and Initial, In Vitro SAR of an Inhibitor of the Shigella Flexneri Virulence Regulator VirF. *PLoS One* **2015**, *10* (9), e0137410. https://doi.org/10.1371/journal.pone.0137410.

(20) Zhao, X.; Li, G.; Liang, S. Several Affinity Tags Commonly Used in Chromatographic Purification. *J. Anal. Methods Chem.* **2013**, *2013* (Table 1). https://doi.org/10.1155/2013/581093.

(21) Koppolu, V.; Osaka, I.; Skredenske, J. M.; Kettle, B.; Hefty, P. S.; Li, J.; Egan, S. M. Small-Molecule Inhibitor of the Shigella Flexneri Master Virulence Regulator VirF. *Infect. Immun.* **2013**, *81* (11), 4200–4207. https://doi.org/10.1128/IAI.00919-13.

(22) Tobe, T.; Yoshikawa, M.; Mizuno, T.; Sasakawa, C. Transcriptional Control of the Invasion Regulatory Gene VirB of Shigella Flexneri: Activation by VirF and Repression by H-NS. *J. Bacteriol.* **1993**, *175* (19), 6142–6149.

(23) Tran, C. N.; Giangrossi, M.; Prosseda, G.; Brandi, A.; Di Martino, M. L.; Colonna, B.; Falconi, M. 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.* **2011**, *39* (18), 8122–8134. https://doi.org/10.1093/nar/gkr521.

(24) Fischer, M.; Leech, A. P.; Hubbard, R. E. Comparative Assessment of Different Histidine-Tags for Immobilization of Protein onto Surface Plasmon Resonance. *Anal. Chem.* **2011**, 1800–1807.

(25) Di Martino, M. L.; Falconi, M.; Micheli, G.; Colonna, B.; Prosseda, G. The Multifaceted Activity of the VirF Regulatory Protein in the Shigella Lifestyle. *Front. Mol. Biosci.* **2016**, *3* (SEP), 1–11. https://doi.org/10.3389/fmolb.2016.00061.

(26) Prosseda, G.; Falconi, M.; Giangrossi, M.; Gualerzi, C. O.; Micheli, G.; Colonna, B. The VirF Promoter in Shigella: More than Just a Curved DNA Stretch. *Mol. Microbiol.* **2004**, *51* (2), 523–537. https://doi.org/10.1046/j.1365-2958.2003.03848.x.

(27) Schneiders, T.; Levy, S. B. MarA-Mediated Transcriptional Repression of the Rob Promoter. *J. Biol. Chem.* **2006**, *281* (15), 10049–10055. https://doi.org/10.1074/jbc.M512097200.

Chapter 3 Molecular Recognition of MarA Binding

AraC family proteins are widespread within bacteria and contain a highly conserved DNA-binding domain (DBD) consisting of two helix-turn-helix (HTH) motifs.[12,13] However, these proteins are often highly insoluble, making them difficult to purify and work with *in vitro*.[27] There have been attempts in the literature and by us (See **Chapter II**) to obtain native VirF to study *in vitro,* but the yields have been extremely low and insufficient for study.[18] Relatively few three-dimensional structures of AraC family proteins have been solved and, unfortunately, no crystal structure of VirF has been reported. This has slowed studies of the VirF-DNA binding interaction thus reducing our ability to improve and develop more potent VirF inhibitors. Structures of a few AraC family proteins from *E. coli* and one from *V. cholerae* have been solved. We selected two proteins, GadX and MarA, for use in homology modeling of the VirF DBD and *in vitro* analyses. GadX is involved in promoting acid resistance and shows a 29% sequence identity to the DBD of VirF.[28] The GadX structure (PDB: 3MKL) does not include DNA and is likely reflective of the GadX conformation in solution.

MarA, a regulator of multiple antibiotic resistance, has been studied extensively and a crystal structure of it bound to its cognate DNA promoter *marRAB* (PDB: 1BL0) has been solved by Rhee et al.[29], who identified amino acids in positions that very likely contribute to the protein-DNA binding interactions via hydrogen bonding or Van der Waals interactions. In 2000, Gillette *et al*.[30] published their results of an *in vivo* alanine scan of 107 amino acids of MarA and evaluated their ability to activate transcription of five *mar* regulon promoters *in vivo*. Six of

those mutants were purified and their *in vitro* affinities for *marRAB* and *micF* were determined in a ^{32}P-labeled DNA binding assay.

Using GadX and MarA as structural models, we prepared two homology models for the VirF DBD, free and DNA-bound. Here, we have conducted our own alanine scan of seven MarA residues, which were identified by Rhee et al. to make base-specific interactions with *marRAB*.[31] Two of these were studied *in vitro* by Gillette et al.[30], we added the other five identified by Rhee et al. and evaluated the affinities of all seven for the *marRAB* promoter DNA *in vitro*. Based on sequence alignments with MarA and the MarA-based VirF homology model, the corresponding residues in VirF were mutated to alanine and tested *in vitro* to both validate the MarA-based model of VirF in a DNA-bound conformation and elucidate key DNA-binding interactions between VirF and the *virB* promoter.

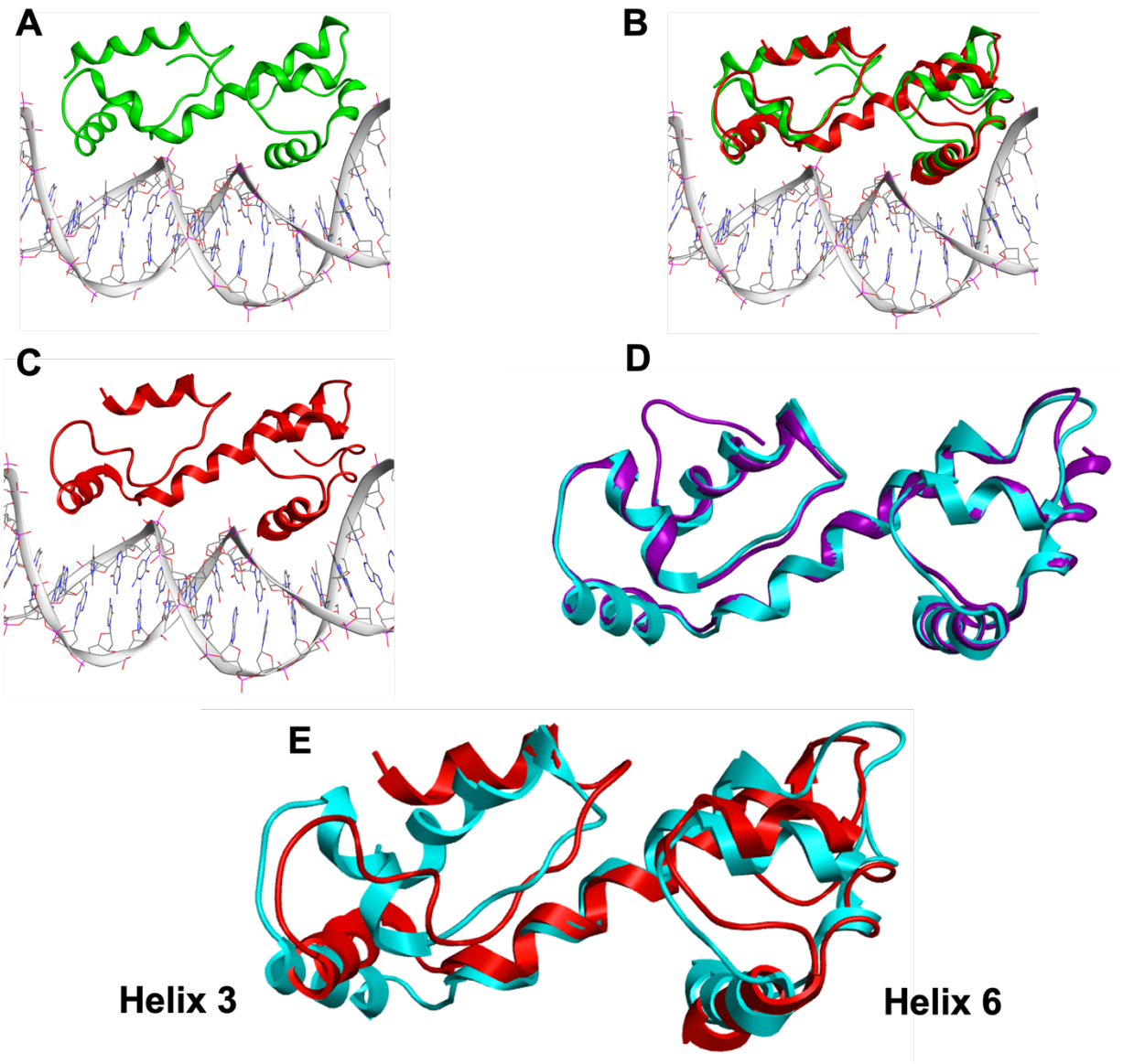


Figure 3-1: ***VirF C-Terminal, DNA Binding Domain (DBD) Homology Models, and the Structural Templates MarA and GadX****. A): Crystal structure of MarA (1BL0; Green) bound to the marRAB promoter DNA. B): The VirF DBD homology model (Red) created using MarA as a template*

Materials and Methods

Reagents

All standard buffer components were purchased from Millipore Sigma or Thermo Fisher. Specific reagents or biological products not purchased from these are noted in parentheses.

DNA oligonucleotides were purchased from Integrated DNA Technologies. Equipment utilized for these experiments was purchased from varying companies which are indicated in parentheses throughout this section.

Alignment and Homology Modeling

Using SWISS-MODEL, the Protein Data Bank database was searched for structures that showed homology with the C-terminal DBD of VirF.[32–36] The two templates selected, GadX (PDB: 3MKL, 2.15 Å resolution) and MarA (PDB: 1BL0, 2.30 Å resolution), stood out at the top of the list with high sequence coverage, identity/homology, and GMQE (Global Model Quality Estimate) scores. To develop the VirF DBD models, the protein sequence was entered into MOE (Molecular Operating Environment). Here, the sequence was used to search the PDB database to obtain the GadX and MarA structures that we wanted to use as templates and both structures were opened for analysis in the program. In the homology model window, the VirF DBD sequence was entered as the sequence to model and each template was selected as a primary structure template in two separate consecutive modeling runs. The geometries of the models with the highest RMSDs for each template were evaluated and selected as the representative homology model. RMSD values were calculated using MOE. The protein sequences were aligned and analyzed using MOE, giving percent-identity match. MarA and VirF DBD were also aligned and evaluated using EXPASY SIM-Alignment Tool.

Alanine-Scanning Mutagenesis

Site-directed mutagenesis of the *marA* and *malE-virF* genes was performed using oligonucleotides described in **Table 3-1**. The alanine mutation was incorporated using the GCG alanine codon for each MalE-VirF mutant (I189, R192, K193, S238, Y239, I241, R242) and each MarA mutant (W42, Q45, R46, Q92, T93, T95, R96) except for R46A, which contains the GCT codon. MalE-VirF numbering refers to the primary sequence of VirF from *Shigella flexneri*. Two-step PCR was used on pET-15b *marA* and pBAD202-MALVirF with PFU Turbo DNA Polymerase in a MiniAmp^{TM} Plus Thermal Cycler (Thermo Fisher). Detailed PCR methods can be found in **Chapter II**. All mutations were confirmed by DNA sequence analysis.

*Table 3-1:**DNA Oligonucleotide Primers Designed and Used for Alanine-Scanning Mutagenesis of the MarA (pET15b-MarA) and MalE-VirF (pBAD202-MALVirF) Expression Plasmids**. Alanine codons incorporated into the expression plasmids are bolded.*

Primer Name	Sequence (5' → 3')
MarA W42A Forward	GTTACTCCAAA**GCG**CACCTGCAACGGATG
MarA W42A Reverse	CATCCGTTGCAGGTG**CGC**TTTGGAGTAAC
MarA Q45A Forward	CCAAATGGCACCTG**GCG**CGGATGTTTAAAAAAGAAACC
MarA Q45A Reverse	GGTTTCTTTTTTAAACATCCG**CGC**CAGGTGCCATTTGG
MarA R46A Forward	TCCAAATGGCACCTGCAA**GCT**ATGTTTAAAAAAGAAACC
MarA R46A Reverse	GGTTTCTTTTTTAAACAT**AGC**TTGCAGGTGCCATTTGGA
MarA Q92A Forward	GATATGGCTTCGAGTCGCAA**GCG**ACTCTGACCCGAACCTTC
MarA Q92A Reverse	GAAGGTTCGGGTCAGAGT**CGC**TTGCGACTCGAAGCCATATC
MarA T93A Forward	GTCGCAACAA**GCG**CTGACCCGAACCTTC
MarA T93A Reverse	GAAGGTTCGGGTCAG**CGC**TTGTTGCGAC
MarA T95A Forward	GAGTCGCAACAAACTCTG**GCG**CGAACCTTCAAAAATTACTTTG
MarA T95A Reverse	CAAAGTAATTTTTGAAGGTTCG**CGC**CAGAGTTTGTTGCGACTC
MarA R96A Forward	GCAACAAACTCTGACC**GCG**ACCTTCAAAAATTACTTTGATGTTCC
MarA R96A Reverse	GGAACATCAAAGTAATTTTTGAAGGT**CGC**GGTCAGAGTTTGTTGC
MalE-VirF I189A Forward	CTTGAATTTATCAGAA**GCG**GCTGTTAGAAAACG
MalE-VirF I189A Reverse	CAATCGTTTTCTAACAGC**CGC**TTCTGATAAATTCAAG
MalE-VirF R192A Forward	GAAATAGCTGTT**GCG**AAACGATTGGAGAGTG
MalE-VirF R192A Reverse	CACTCTCCAATCGTTT**CGC**AACAGCTATTTC
MalE-VirF K193A Forward	CAGAAATAGCTGTTAGA**GCG**CGATTGGAGAGTG
MalE-VirF K193A Reverse	CACTCTCCAATCG**CGC**TCTAACAGCTATTTCTG
MalE-VirF S238A Forward	GGAATATCAAGCCCA**GCG**TATTTTATAAGGAAATTT
MalE-VirF S238A Reverse	AAATTTCCTTATAAAATA**CGC**TGGGCTTGATATTCC
MalE-VirF Y239A Forward	GGAATATCAAGCCCATCT**GCG**TTTATAAGGAAATTT
MalE-VirF Y239A Reverse	AAATTTCCTTATAAA**CGC**AGATGGGCTTGATATTCC
MalE-VirF I241A Forward	GCCCATCTTATTTT**GCG**AGGAAATTTAATGAAT
MalE-VirF I241A Reverse	ATTCATTAAATTTCCT**CGC**AAAATAAGATGGGC
MalE-VirF R242A Forward	GCCCATCTTATTTTATA**GCG**AAATTTAATGAATATTATGGT
MalE-VirF R242A Reverse	ACCATAATATTCATTAAATTT**CGC**TATAAAATAAGATGGGC

MarA Expression and Purification

Our procedure to express and purify MarA was adapted from Jair *et al*.[37] Starter cultures (10 mL) of *E. coli* BL21(DE3) containing pET15b-*marA* were grown overnight in 2xTY media broth (16 g bactotryptone, 10 g yeast extract, 5 g NaCl per liter of water) supplemented with carbenicillin at 37 °C under vigorous agitation. The next day, the starter culture was used to inoculate 1 L of 2xTY broth supplemented with carbenicillin. The cells were grown to an OD_{600} = 0.8 before expression was induced with the addition of isopropyl β-D-1-thiogalactopyranoside

(IPTG) at a final concentration of 0.4 mM and the culture continued to shake overnight at 16 °C. Cells were then harvested by centrifugation (6000 x g, 4 °C, 15 min) before being resuspended in 25 mL MarA lysis buffer (50 mM Tris-HCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 M NaCl, pH 7.5) supplemented with a "c0mplete, Mini Protease Inhibitor Cocktail" tablet (Roche) and 0.1 mM phenylmethylsulphonyl fluoride (PMSF). All proceeding steps were performed on ice or at 4 °C. Cells were lysed via sonication (8 cycles, 15 sec pulse, 3 min intervals, 60% setting) utilizing an ultrasonic XL2020 sonicator (Misonix). Following sonication, the solution was pelleted by ultracentrifugation (120,000 x g, 4 °C, 30 min). The supernatant was discarded, and the pellet was washed with 30 mL of MarA denature buffer 1 (50 mM Tris-HCl, 4 M urea, pH 8.5) before repeating the ultracentrifugation. The supernatant was discarded again, and the pellet was resuspended in 25 mL MarA denature buffer 2 (50 mM Tris-HCl, 6 M guanidinium chloride, pH 8.5). The mixture was subjected to ultracentrifugation a third time, collecting the supernatant. Next, 2 mL of Ni-NTA agarose (Qiagen) were added, and the solution was rocked gently overnight at 4 °C. The following day, the resin slurry was poured into an empty purification column. Once settled, the resin bed was washed stepwise with increasing concentrations of imidazole by combining MarA elution buffer (50 mM Tris-HCl, 500 mM NaCl, 1 M imidazole, pH 8.5) and MarA wash buffer (50 mM Tris-HCl, 500 mM NaCl, pH 8.5) to 100 mM, 300 mM, 500 mM, and 750 mM imidazole. The fraction eluted with 300 mM imidazole contained MarA, verified by SDS-PAGE, and was dialyzed with a Slide-A-Lyzer dialysis cassette (3K MWCO; Thermo Fisher) at 4 °C overnight in MarA dialysis buffer (50 mM Tris-HCl, 250 mM NaCl, 0.1% Triton X-100, 20% glycerol, pH 8.0). The protein concentrations were determined with a Bradford assay (BioRad) using bovine serum albumin (BSA) standards. Protein stocks were flash frozen in liquid nitrogen and then stored at -80 °C.

MalE-VirF Expression and Purification

The expression and purification of MalE-VirF was carried out by my colleague Garrett Dow. WT and mutant MalE-VirF proteins were expressed similarly to that previously described in Emanuele and Garcia.[24] TOP10 *E. coli* were transformed with the MalE-VirF expression plasmid, pBAD202-MALVirF. TOP10 cells harboring the expression plasmid, and supplemented with kanamycin, were grown and protein expression was induced as previously described[24]. The cells were resuspended in MalE-VirF binding buffer (20 mM Tris-HCl, 1 mM EDTA, 1 mM dithiothreitol (DTT), 500 mM NaCl, pH 7.4) and supplemented with a "c0mplete, Mini Protease Inhibitor Cocktail" tablet (Roche) and 0.1 mM PMSF. Following resuspension, all subsequent steps were on ice or kept at 4 ˚C. The resuspended cells were lysed via sonication (6 cycles, 30 seconds pulse, 4 minutes rest, 60% of max pulse setting) using an ultrasonic XL2020 sonicator (Misonix). Cellular debris was separated via centrifugation (45 min, 25000xg, 4 ˚C). Following centrifugation, the supernatant was loaded onto a 5 mL MBPTrap HP column (Cytiva) using an AKTA FPLC (GE Healthcare). The column was then washed with 15 column volumes (CV) MalE-VirF binding buffer. Protein was eluted from the column using MalE-VirF elution buffer (20 mM Tris-HCl, 1 mM EDTA, 1 mM DTT, 500 mM NaCl, 5% glycerol, 10 mM maltose, pH 7.4). Eluent was collected in 1.5 mL fractions. Fractions with the highest resulting UV-280nm absorbance were collected and concentrated to approximately 600 µL using Amicon Ultra-15 MWCO 10 kDa centrifugal filter units and then filtered to remove any precipitate. The resulting filtered protein was loaded onto a Superdex 200 GL10/300 gel filtration column to separate MalE-VirF from truncated MalE and other protein impurities eluted from the MBPTrap HP column. Protein was eluted from the column using MalE-VirF binding buffer and 0.5 mL

fractions were collected and tested via SDS-PAGE to determine where MalE-VirF eluted. The corresponding MalE-VirF fractions were dialyzed with a Slide-A-Lyzer dialysis cassette (10K MWCO; Thermo Fisher) at 4 ˚C overnight in MalE-VirF dialysis buffer (20 mM Tris-HCl, 1 mM EDTA, 5 mM DTT, 200 mM NaCl, 40% glycerol, pH 7.4). The concentration of each protein was tested with a Bradford assay (BioRad) using BSA standards. Protein stocks were flash frozen in liquid nitrogen and then stored at -80 ˚C.

Electrophoretic Mobility Shift Assay (EMSA)

EMSAs were performed as previously described in Emanuele and Garcia, 2015.[24] Prior to preparing the gel and reactions, the Cy5-labeled *virB* (*pvirB*) and *marRAB* oligonucleotide promoter probes were annealed using the oligonucleotide primers found in **Table 3-2.**[24] The *marRAB* probe contains the MarA binding site at the center of the 60 bp DNA whereas *pvirB*, also 60 bp, contains two VirF binding sites with extra flanking bases to support DNA duplex stability.[14,24,31] Although in the crystal structure, MarA is bound to a 20 bp DNA fragment, the length of our probe was increased to support duplex stability and for consistency with previous MalE-VirF EMSAs.[24,31] A 6% native polyacrylamide gel was prepared using 30% acrylamide/bis-acrylamide (29:1 ratio) solution and TBE buffer (0.25x final concentration; 22 mM Tris Base, 22 mM boric acid, 0.5 mM EDTA, pH 8.5 or 9.5). EMSAs containing MarA proteins and the *marRAB* promoter used 0.25x TBE gel and running buffer at pH 8.5 whereas EMSAs containing MalE-VirF proteins and *pvirB* were run with the same buffer at pH 9.5. MalE-VirF was not observed to enter the native polyacrylamide gel when the 0.25x TBE buffer was below pH 9.5. MarA and MalE-VirF WT and mutant protein concentrations varied by purification yield.

For MarA EMSAs, each reaction was composed of 3 µL of *marRAB* probe (42 nM), 9 µL MarA WT or mutant protein, 1 µL salmon sperm DNA (Invitrogen; 0.7 mg/mL), 0.5 µL BSA (0.07 mg/mL), and 1.5 µL Milli-Q H_2O for a final volume of 15 µL. Final concentrations of DNA probe, salmon sperm DNA, and BSA within each reaction are presented in parentheses. Final buffer conditions were 40 mM Tris-HCl, 160 mM NaCl, 0.04 mM EDTA, 0.06% Triton X-100, 12% glycerol, pH 8.0. Serial dilutions were prepared by diluting the highest possible concentration for each protein with MarA dialysis buffer. Additionally, we ordered and obtained a 143 bp 5'-Cy5 labeled *marRAB* promoter from IDT which was used to compare our assay protocol and DNA promoter with those utilized in Gillette et al.[30]

For testing with MalE-VirF, reactions were prepared with 6 µL MalE-VirF WT or mutant protein, 6 µL *pvirB* (83 nM),1.5 µL Milli-Q H_2O, 1 µL salmon sperm DNA (0.7 mg/mL), and 0.5 µL BSA (0.07 mg/mL). Final buffer conditions were 12 mM Tris-HCl, 100 mM NaCl, 16% glycerol, 2 mM DTT, and 0.44 mM EDTA, pH 7.4. To perform titrations of all MalE-VirF proteins against *pvirB*, two-fold dilutions were prepared using MalE-VirF dialysis buffer. Negative control reactions were prepared with 6 or 9 µL native gel loading dye (300 mM Tris-HCl, 50% glycerol, 0.05% bromophenol blue, pH 7) instead of MalE-VirF or MarA proteins, respectively. Prior to loading the reaction mixtures, the empty gel was electrophoresed for 1 hour at 150 V in 0.25x TBE. All reactions were incubated at 37 ˚C for 15 min in a water bath before adding 6 µL of each reaction to the corresponding wells on the gel. The gel was electrophoresed at 150 V for an additional 1.5 hours in the dark at 4 ˚C. Gel visualization was performed using a Molecular Dynamics Typhoon 9200 Molecular Imager by excitation (Ex) at 607 nm and reading the 710 nm emission (Em). Quantitative data was obtained by measuring

the density of the bands on the gel using ImageJ software[38]. Prism 9 software[39] (one-site specific binding equation) was used to generate binding affinities (K_D) and B_{Max} (%).

*Table 3-2**: DNA Oligonucleotides Used for Preparation of EMSA and FP Assays**. Underlined sequences represent the individual binding sites on the marRAB and virB promoters, and the italicized sequence represents the LUEGO (labeled universal gel shift oligonucleotide*

Primer Name	Sequence (5' → 3')
marRAB Top Strand EMSA	CATTGAACAAAACTTGAACCGATTTAGCAAAACGTGGCATCGGTCAA TTCATTCATTTGA (5'-6FAM used in FP)
marRAB Bottom Strand EMSA	TCAAATGAATGAATTGACCGATGCCACGTTTTGCTAAATCGGTTCAA GTTTTGTTCAATG*CCAGACCAGGGCAC*
marRAB Bottom Strand FP	CATTGAACAAAACTTGAACCGATTTAGCAAAACGTGGCATCGGTCAA TTCATTCATTTGA
marRAB Top Strand Short FP	GCTGGCTACGGTGCAAAACGATTTAGCCAAGC (either a 5'-6FAM, -Cy3, or -Cy5)
marRAB Bottom Strand Short FP	GCTTGGCTAAATCGTTTTGCACCGTAGCCAGC
LUEGO	Cy5-GTGCCCTGGTCTGG
pvirB Top Strand FP and EMSA	AGAATATTATTCTTTTATCCAATAAAGATAAATTGCATCAATCCAGCTA TTAAAATAGTA (5'-6FAM used in FP)
pvirB Bottom Strand EMSA	TACTATTTTAATAGCTGGATTGATGCAATTTATCTTTATTGGATAAAAG AATAATATTCT*CCAGACCAGGGCAC*
pvirB Bottom Strand FP	TACTATTTTAATAGCTGGATTGATGCAATTTATCTTTATTGGATAAAAG AATAATATTCT
pvirB Top Strand Scram1 EMSA	AGAATAATTAATCTTCTATTCTTAAAGATAAATTGCATCAATCCAGCTA TTAAAATAGTA
pvirB Bottom Strand Scram1 EMSA	TACTATTTTAATAGCTGGATTGATGCAATTTATCTTTAAGAATAGAAGA TTAATTATTCT*CCAGACCAGGGCAC*
pvirB Top Strand Scram2 EMSA	AGAATATTATTCTTTTATCCAATAAAGATAAATTGCATACGATATCAAC AATCTATAGTA
pvirB Bottom Strand Scram2 EMSA	TACTATAGATTGTTGATATCGTATGCAATTTATCTTTATTGGATAAAAG AATAATATTCT*CCAGACCAGGGCAC*

Fluorescence Polarization (FP) Assay

The FP assays were performed as described in Emanuele and Garcia.[24] The assays were conducted in low-volume round-bottom 384-well plates (Corning). The fluorophore-labeled *pvirB* and *marRAB* oligonucleotide probes were annealed as previously described using the oligonucleotide primers found in **Table 3-2.**[24] The top strands of the *pvirB* or *marRAB* promoters (**Table 3-2**) contained either a 5'-fluorescein, 5'-Cy3, or 5'-Cy5, and were annealed to their corresponding, unlabeled bottom strands to prepare stocks with concentrations of 5 μM. The *pvirB* probe contained a 5'-fluorescein whereas the *marRAB* promoter probes were prepared with all three fluorophores, individually. The DNA probes were diluted with probe buffer (50

mM Tris-HCl, 1 mM EDTA, 5 mM DTT, and 200 mM NaCl, pH 7.4) to a final working concentration of 20 nM and supplemented with 1.4 mg/mL salmon sperm DNA (Invitrogen) and 0.14 mg/mL BSA. Next, 14 two-fold serial dilutions of MarA WT or MarA mutant were prepared with MarA dialysis buffer then 10 µL of each dilution were added to the appropriate wells in triplicate for each concentration tested. For MalE-VirF, two-fold dilutions were prepared using MalE-VirF dialysis buffer prior to loading 10 µL of each dilution onto the plate. Following loading of protein into the wells, 10 µL of the 20 nM DNA probe solution were added to each well. Final concentrations of DNA probe, salmon sperm DNA, and BSA in each reaction are 10 nM, 0.7 mg/mL, and 0.07 mg/mL, respectively. Negative control reactions were also tested in triplicate, containing 10 µL MarA or MalE-VirF dialysis buffers and 10 µL of their corresponding 20 nM DNA probe solutions to determine the baseline FP for each DNA probe. Additionally, for each protein concentration tested, a blank reaction, containing the corresponding tested protein concentration and the DNA probe solution lacking the labeled probe, was used to subtract fluorescent contributions of MalE-VirF or MarA, salmon sperm DNA, BSA, and other buffer components from the test reactions. The plate was incubated at 37 °C for 2 hours before the raw fluorescence was measured using a Biotek Synergy H1 plate reader after excitation at the appropriate wavelength for the corresponding *pvirB* and *marRAB* probes (fluorescein Ex/Em = 485/528, Cy3 Ex/Em = 554/568, Cy5 Ex/Em = 649/666). Fluorescence polarization (FP) was calculated from the corrected parallel ($F_{\parallel}$) and perpendicular ($F_{\perp}$) fluorescence values using **Equation 1**. The G factor, often used in polarization calculations, was not included in our calculations as we determined it to be negligible under these conditions (data not shown).

Equation 1 $FP = (F_{\parallel} - F_{\perp}) / (F_{\parallel} + F_{\perp}) * 1000$

The plots were fit by non-linear regression to the following sigmoidal four-parameter equation using Prism 9[40] (GraphPad Software; **Equation 2**),

Equation 2 $mP = min + (max - min) / (1 + 10^{((logEC_{50}-X) * Hill\ Slope)})$

where max and min were the maximum and minimum plateaus of the mP and X is the log of sample concentration. When unconstrained, the values of max, min, EC_{50}, and Hill slope are fit by the regression plot. For all plots except the I189A mutant, the mP max was constrained to the observed mP max for WT MalE-VirF (e.g., 95). The mP plot for the I189A mutant is shifted ~10 mP units higher than the other plots, therefore the mP max was fit. The span (mP range) for the I189A mutant matched that of the WT and other mutants.

Circular Dichroism (CD)

Buffer exchange was performed on WT and selected MalE-VirF and MarA mutant proteins to place the samples in 10 mM NaH_2PO_4 (pH 7.5) using Amicon Ultra-0.5 centrifugal filter units (3 MWCO; MilliporeSigma). The samples were centrifuged (13,000xg, 4 ˚C, 10 min), flow-through was discarded, and then diluted again with 10 mM NaH_2PO_4 (pH 7.5). This process was repeated for 5-10 rounds. The concentration of each sample was determined via Bradford assay (BioRad). The resultant samples were flash frozen in liquid nitrogen and stored at -80 ˚C prior to CD testing. For CD testing, samples were loaded into a 1 mm path-length quartz cuvette and CD spectra were collected from 195-250 nm using a JASCO J-810

spectrometer. Spectra of a sample containing only buffer was used to correct the raw data for each protein. JASCO Spectra Manager was used to visualize the spectra and export the raw data to be plotted using Prism 9 software.[41]

Results

MarA vs VirF: Sequence and Structural Homology Models

The VirF DBD was subjected to a homology search using SWISS-MODEL to identify suitable structural templates.[34] GadX (3MKL) had a sequence identity score of 29 and a Global Model Quality Estimation (GMQE) score of 0.76. MarA (1BL0) had a sequence identity score of 19 and a GMQE score of 0.70. As discussed above, these two structures were chosen as templates for our homology modeling.

The qualities of the generated homology models **(Figure 3-1)** were assessed by evaluating the stereochemical parameters on a Ramachandran plot using MOE **(Figure 3-2).** In the GadX-based homology model, there are no bond angles that form sterically disallowed confirmations. Two residues fall just outside of the ideal tertiary structure regions (outlined in blue in **Figure 3-2 A**), but still within the sterically allowed region for beta sheets (within the red outline). The MarA-based homology model is similar, but with more residues falling into the sterically allowed region of the plot for atoms that are closer together (outside the blue line but within the red in **Figure 3-2 B**). This model also contains one outlier, Serine 38, which lies outside of the sterically allowed region **(Figure 3-2 B).**

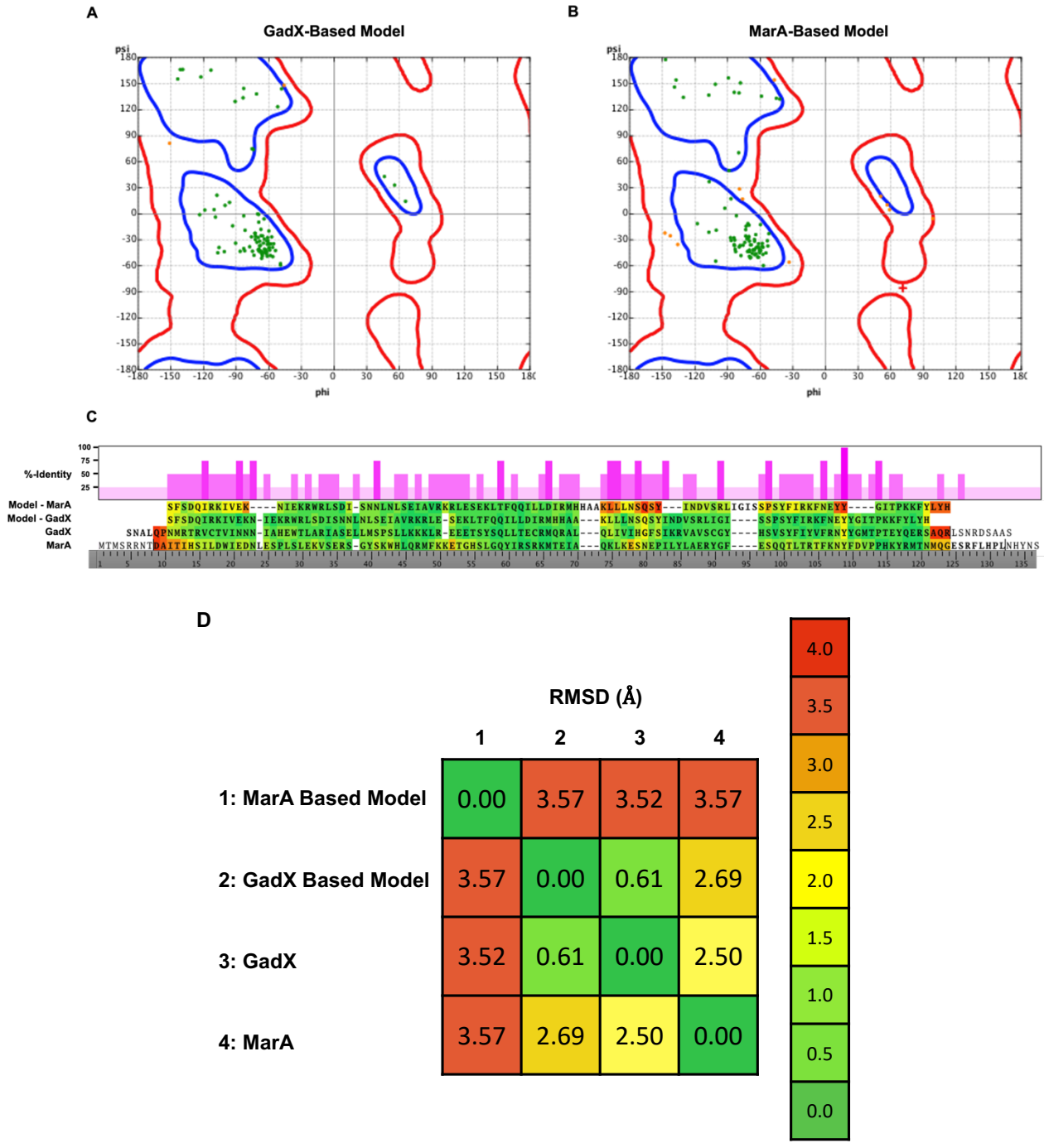


Figure 3-2: ***Homology Model Validation****. Ramachandran plots of the A): GadX-Based and B): MarA-based VirF homology models were created using MOE. Green points fall within the blue outlined region, indicating that there are no steric clashes in the beta-sheet confirmations (top left blue outline), right-handed alpha-helical confirmations (middle left blue outline), or left-handed alpha-helical confirmations (middle right blue outline). Orange points fall outside of the blue outline but within the red outline, indicating they have closer but still sterically allowed confirmations. Red points are outliers that are in sterically disallowed confirmations. C): Sequence alignments of both homology models and the two template structures. Percent identity is displayed in the pink bars above the sequence and RMSD is depicted by color highlighting the sequence and RMSD is depicted by color highlighting the sequence shown in D. D): RMSD table comparing each of the homology model with GadX and MarA.*

When superimposing the VirF models with the templates **(Figure 3-1 B and D)**, the percent identity is plotted above the sequence alignment and the Root Mean Square Deviation (RMSD) for each residue is color coded on the sequence (spectrum from green to red) **(Figure 3-2 C)**. Several regions of the proteins have high identity and a small RMSD value, particularly the GadX-based model. In the MarA•*marRAB* structure, helix-3 and helix-6 are binding to the major groove on the DNA promoter (**Figure 3-1 A and Figure 3-3**). Helix-6 shared a similar position in space among the two homology models and two crystal structures, where helix-3 showed variability in its placement, though the percent identity did not vary much between the two regions. In the MarA-based VirF model, helix-2 is shown as a free loop, but when overlayed with MarA or the GadX-based model (**Figure 3-1 B and E**), it appears that this region has helical characteristics. This is likely an artifact of MarA having lower sequence identity to the VirF DBD. We decided to use the model that MOE created and chose not to impose this region into an alpha helix. The alignment of the homology model with MarA allowed us to visualize which side chains on these binding regions (helix-3 and helix-6) could potentially be contributing to VirF's ability to bind to *pvirB* **(Figure 3-1 and Figure 3-3)**.

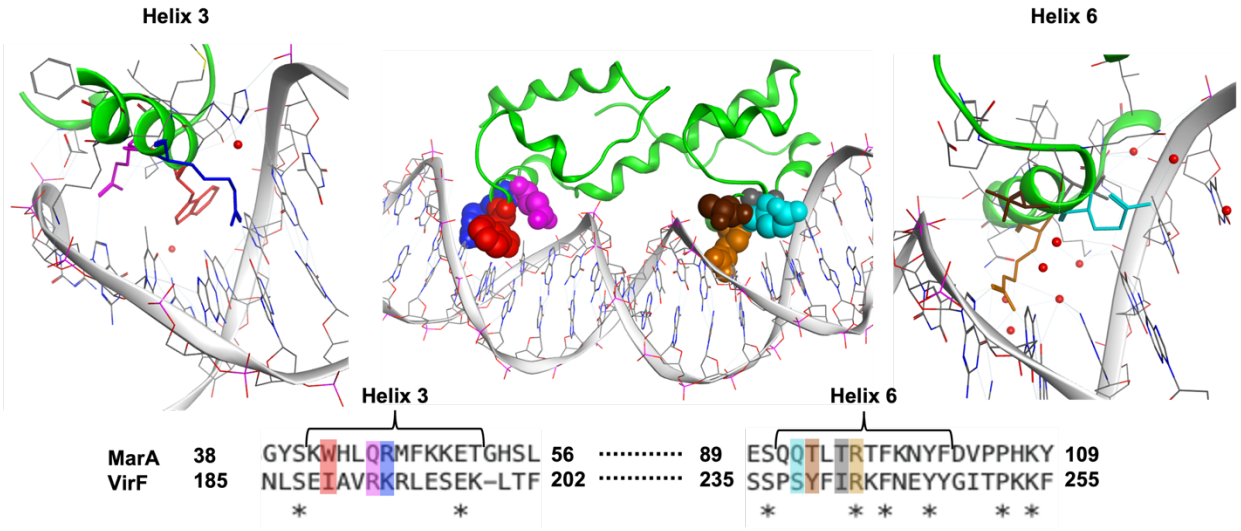


Figure 3-3: ***Spatial Orientation of the Seven Amino Acids in MarA (PDB ID: 1BL0) That Make Base Contacts with the marRAB Promoter.*** *Residues are color coded according to the alignment of the MarA and VirF DNA-binding domains supplemented below. Small red spheres are water molecules.*

Purification of MarA and the DNA Binding Domain Mutants

The purification of active wild type MarA was achieved based on the published protocol by Jair et al.[37] Most published protocols to obtain the protein purify it from the insoluble pellet; however, we had difficulties isolating protein that was active in our DNA binding assays. We attempted to replicate the procedure of Gillette et al.[30], subjecting the cell pellet to high concentrations of guanidinium HCl before lysing via dry ice bath. Unfortunately, MarA isolated this way showed no shift in the EMSA that would indicate DNA binding. When we replicated the Jair et al.[37] procedure as published, we run into the same issue. MarA that is isolated through this method showed no binding to *marRAB* in the EMSA. To finally obtain active MarA, we followed this method until the end of third spin in the ultracentrifuge. After this step, we decided not to dilute our supernatant with tris buffer before applying the solution to the column. Instead, we sterile filtered the supernatant and added 2 mL of Ni-NTA agarose (Qiagen) and rocked the solution gently at 4°C overnight. The resin slurry was poured into an empty column and once

settled, the resin bed was washed stepwise with increasing concentrations of imidazole. The fraction eluting at 300 mM imidazole typically contained MarA **(Figure 3-4 A)**. The activity of this MarA preparation was confirmed by EMSA, probing its ability to bind to our *marRAB* probe and its specificity via its inability to bind to our *pvirB* probe at two different concentrations **(Figure 3-4 B)**. Once WT MarA was successfully isolated and activity was observed, we used this method to express and purify the rest of the DBD mutants (W42A, Q45A, R46A, Q92A, T93A, T95A, R96A).

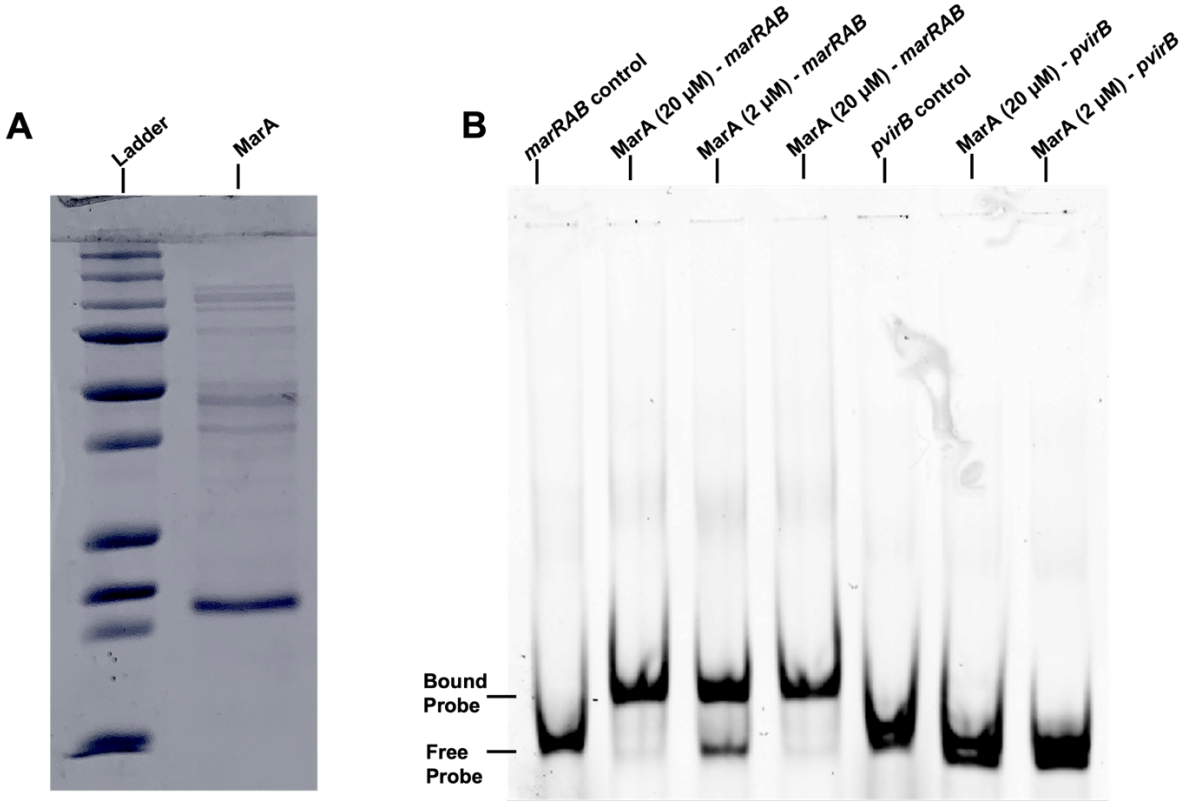


*Figure 3-4: **MarA WT Purification Gels**. A): SDS PAGE post MarA purification. Fractions representing distinct steps during the process could not be visualized due to the high concentrations of guanidinium HCl and urea. B): EMSA gel displaying MarA's ability to bind the marRAB promoter at 2 and 20 µM. We also demonstrated that MarA wt could not recognize and bind the pvirB probe.*

MarA DNA Binding

DNA-binding assays were used to identify the binding activity of MarA for their cognate promoters (promoter sequences, fluorophores used, and method to obtain probes can be found in **Table 3-2** and **Figure 3-5**). First, to evaluate the ability of MarA to bind to its DNA promoter, alanine scanning mutagenesis was utilized, creating seven MarA mutants. Three mutations were made to helix-3 in the first HTH motif (W42A, Q45A, and R46A) and four were made to helix-6 in the second HTH (Q92A, T93A, T95A, and R96A). To evaluate how these alanine mutations affected binding, an EMSA was used to compare each mutant to the WT MarA (**Figure 3-6**). The quantitative analyses to obtain the binding affinities for all MarA proteins for *marRAB* is presented in **Figure 3-7**. Three of the mutants, W42A, R46A and R96A, did not exhibit saturation of the DNA at any concentration tested. When their data were plotted and fit, none of them achieved a B_{Max} of > 40%, indicating that these mutations dramatically impaired DNA binding. Consequently, the fitted parameters for these three mutants should be considered as estimates. For R46A, the two highest concentration points were plotted, but not used in the non-linear regression. The helix-3 mutant Q45A slightly attenuated MarA's ability to bind to the *marRAB* promoter, with an experimental K_D value of 8.1 μM, relative to 4.8 μM for WT MarA. Interestingly, R46A exhibited significant binding to *marRAB* at concentrations < 3.5 μM (**Figure 3-6**), but the band shift is weaker at concentrations higher than 3.5 μM (**Figure 3-7**). This suggests that the mutant protein may be aggregating at higher concentrations. The only mutation on helix-6 that attenuated binding was R96A. This change significantly increased the experimental estimated K_D to 20 μM. The other three helix-6 mutants all slightly strengthened MarA's affinity for *marRAB*, specifically Q92A, which presented a K_D of 0.23 μM. Both T93A and T95A improved binding to a lesser degree, displaying K_D's of 2.0 μM and 2.5 μM, respectively. Circular dichroism (CD) spectra were collected for WT and our weakest binding

mutant, W42A, to determine if the W42A protein was folded properly (data not shown). W42A presented a similar CD spectrum to MarA WT indicating it is likely folded properly. Since the other mutants exhibited binding in the EMSA, we surmise that all proteins were properly folded.

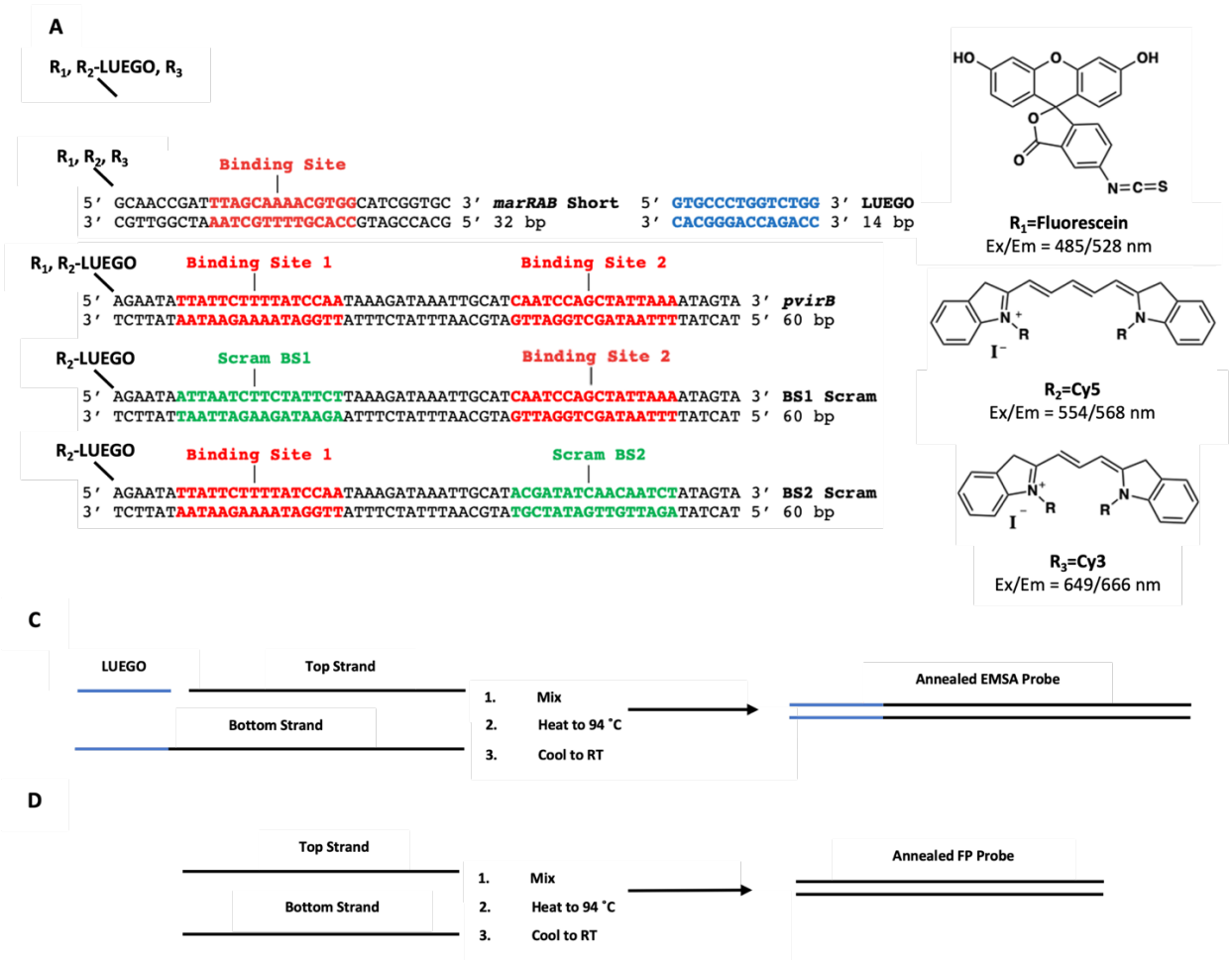


Figure 3-5: ***EMSA and FP Probes for the DNA Promoters Tested in this Report.*** *A): DNA sequences for pvirB and marRAB sequences including protein binding sites labeled in red (green for scrambled binding sites). B): Fluorophores used in the FP and EMSA assays which were placed at the 5' end of the top strands of all probes (EMSA probes contain a 5'-Cy5 label on the LUEGO sequence which resides on the 5' end of the promoter sequence, i.e., R_2-LUEGO). Probes were prepared via the above protocols in C) and D) for the EMSA and FP, respectively (see* ***Table 3-2*** *for oligonucleotide sequences).*

Fluorescence polarization (FP) experiments were attempted as an additional method to assess the binding of each of our MarA mutants. We first tried a 5'-6FAM fluorophore on a 60 bp probe of the *marRAB* DNA (**Table 3-2**). Unfortunately, the data showed no trends in polarization when the protein was added at any concentration. We reasoned that the size

difference between MarA and the large DNA probe was too small to adequately determine polarization if binding was occurring. To remedy this, we reduced the size of the probe from 60 bp to 32 bp while keeping the binding site intact (**Figure 3-5**). Concerns about the fluorescence half-life led us to make three of these shorter probes, each with a different 5' fluorophore (-6FAM, -Cy3, or -Cy5). Regrettably, the same issue was observed when using each of these new probes in the FP experiments. The polarization data showed no determinable trend while titrating in MarA WT, MarA W42A, or MarA T95A (data not shown).

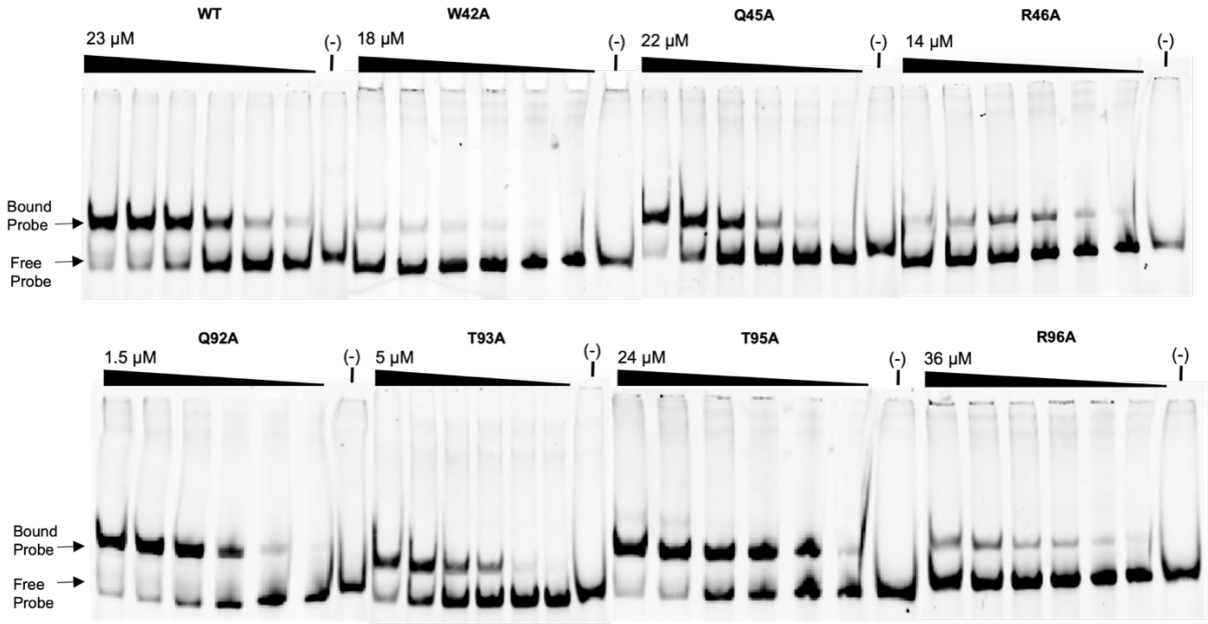
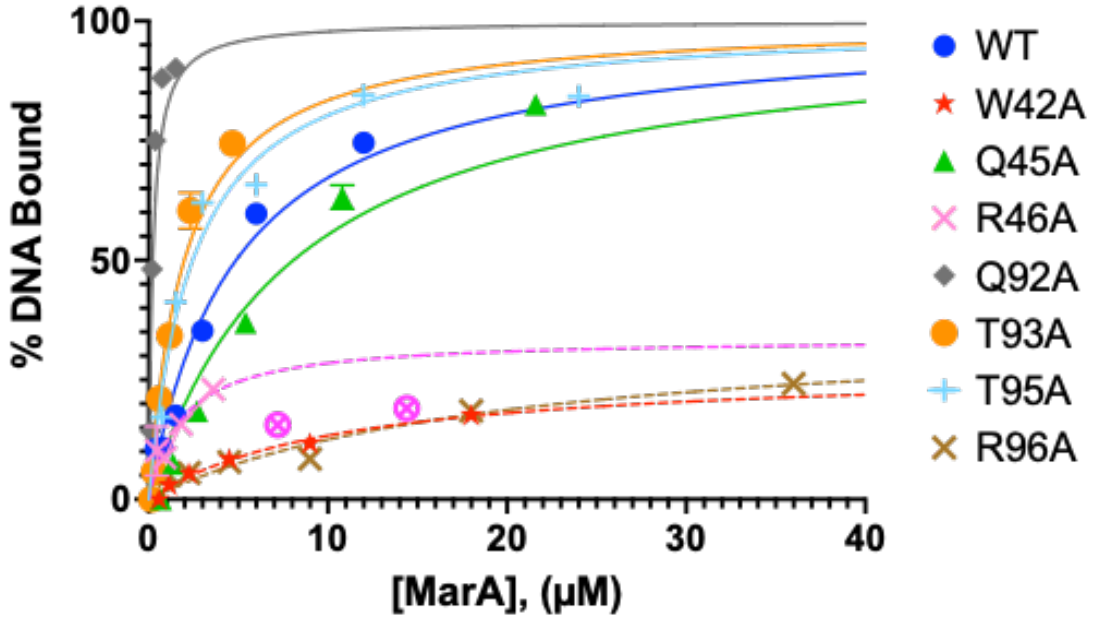


Figure 3-6: ***EMSAs Displaying MarA WT and Each of the MarA Alanine-Mutants' Ability to Recognize and Bind the marRAB Promoter.*** *Each protein was tested in two-fold serial dilutions from the starting concentration shown on the left of each gel*



Protein	WT	W42A†	Q45A	R46A†	Q92A	T93A	T95A	R96A†
K_D, μM	4.8	*11*	8.1	*1.9*§	0.23	2.0	2.5	*20*
(95%CI)	(2.9 - 6.5)	*(6.6 - 20)*	(6.1-10)	*(0.40-12)*	(0.14 - 0.33)	(1.5 - 2.4)	(2.0 – 3.0)	*(9.2 - 49)*
B_{Max}, %	100	*28*	100	*34*	100	100	100	*37*
(95%CI)	-	*(22 - 38)*	-	*(20-ND)*	(88 - ND)	(90 - ND)	-	*(26 - 62)*

Figure 3-7: ***Non-linear Regression Analysis of MarA Alanine-Mutants Compared to MarA WT by EMSA.*** *Binding affinities are presented in the table with 95 % confidence intervals presented in parentheses (ND = not determined). Fits were constrained to B_{Max} = 100 except for W42A, R46A, and R96A which did not reach full saturation. Every protein concentration data point is tested in duplicate and error bars are included indicating the standard deviation. § The two R46A data points circled were not included in the fit. †For W42A, R46A and R96A, the fits are displayed as dashed lines and the fitted parameters (in italics) are included as estimates only since the proteins were unable to fully saturate marRAB in the EMSA experiments*

Previously, our lab screened ~140,000 compounds against VirF in a *Shigella*-based β-galactosidase reporter assay and discovered five lead compounds that displayed IC_{50} values less than 100 μM.[22,23] Of these compounds, 19615 was shown to inhibit VirF DNA binding activity for the *virB* promoter in an EMSA.[24] Due to the high structural and sequence homology of the DNA binding domain of AraC family members (**Figure 1-2**), the specificity of the inhibitor was evaluated. An EMSA was conducted that tested the capability of 19615 to inhibit the MarA-

marRAB interaction in WT MarA as well as each of the DBD mutants **(Figure 3-8)**. At 100 µM inhibitor, no reduction in binding was observed for any of the mutants tested.

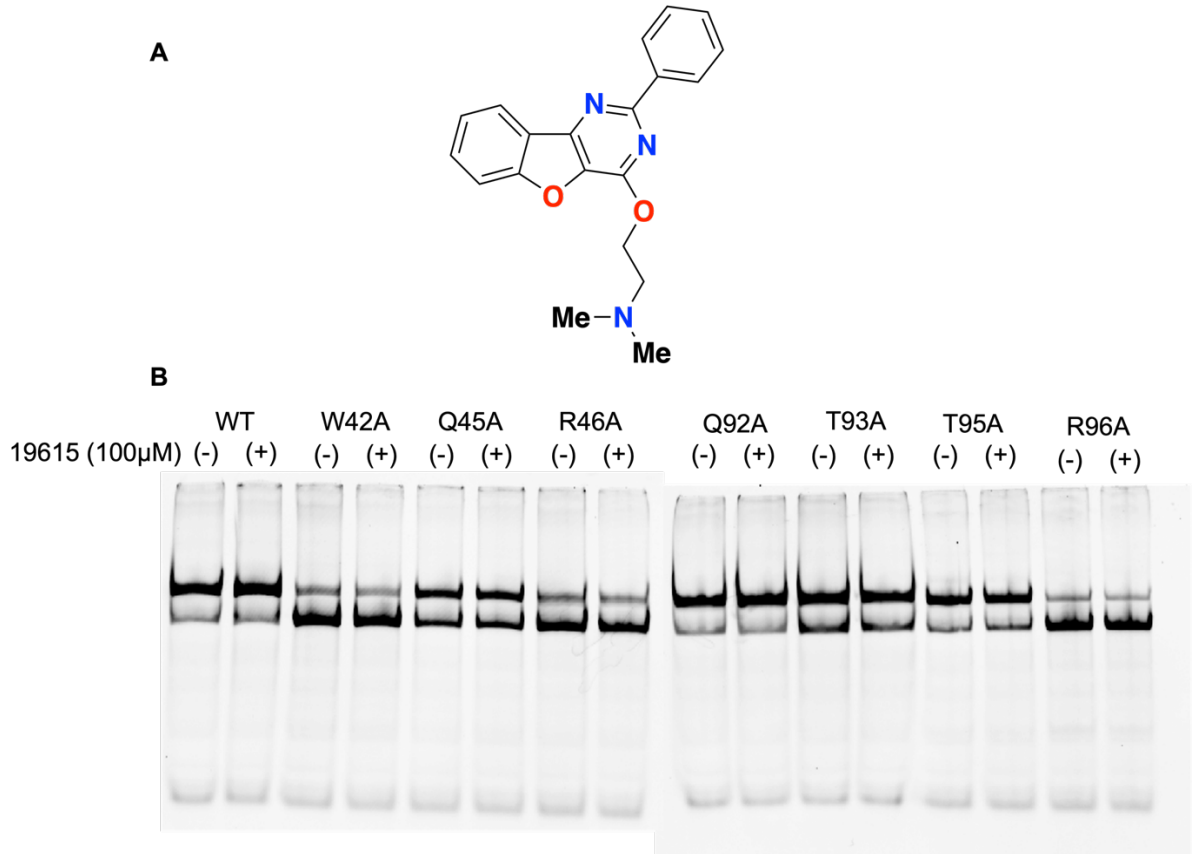


Figure 3-8: ***19615 Structure and Activity Against MarA DBD Mutants.*** *A): Structure of 19615. B): EMSA gel showing each mutant with (+) or without (-) inhibitor present.*

Simultaneous to the above experiments, another lab member, Garrett Dow, conducted the expression and purification of MalE-VirF WT and mutants and evaluated their activity by EMSA and FP. Those results compared to MarA are displayed in **Table 3-3.**

Table 3-3: ***Comparing the Binding Affinities of WT and Mutant MarA and MalE-VirF Proteins for Their Cognate Promoters.*** *Relative K_D's compare each mutation to their WT counterpart. (NDA = no detectable activity, * aggregation was inferred due to very steep Hill slope in FP experiment (data not shown), § estimated K_D determined from* ***Figure 3-7****).*

MarA Mutation	K_D, μM (Determined by EMSA)	Relative K_D	MalE-VirF Mutation	K_D, μM (Determined by FP)	Relative K_D
Wild Type	4.8	1	Wild Type	2.3	1
W42A	11§	2.3	I189A	4.2	1.8
Q45A	8.1	1.7	R192A	NDA	-
R46A	1.9§	0.40	K193A	NDA	-
Q92A	0.23	0.04	S238A	6.3	2.7
T93A	2.0	0.42	Y239A	11*	4.8
T95A	2.5	0.57	I241A	6.5	2.8
R96A	20§	4.1	R242A	10.1	4.4

Discussion

Homology modeling is a powerful tool when experimental determination of the three-dimensional structure of a protein (or other large biomolecule) is not feasible, and structures of homologous proteins are available. Although many attempts have been made, there are no experimental structures of full length VirF or the VirF DBD. VirF is a member of the AraC family of transcriptional regulators, which is widely distributed across gram-negative bacteria and is typically involved in carbon metabolism, stress response, and virulence. Members of this family have a highly conserved C-terminal domain that contains two helix-turn-helix (HTH) DNA binding motifs[12,13]. This universally conserved region of the protein family allowed us to find several templates in the protein data bank with high sequence coverage and strong GMQE scores to the VirF DBD (amino acids 144-262 in WT VirF).

The two structures selected to serve as templates for homology modeling were GadX (PDB ID: 3MKL) and MarA (PDB ID: 1BL0). In the GadX structure, the protein is not bound to DNA. In the MarA structure, it is bound to the *marRAB* promoter (**Figure 3-1 A**). When superimposed, the MarA-based and GadX-based VirF homology models have slightly different orientations of the HTH DNA-binding domains (**Figure 3-1 E**). This could potentially represent

bound and unbound conformations of the Arac-family DBDs, but further analysis is needed to be certain. By overlaying the MarA-based VirF DBD homology model with the MarA crystal structure, not only do we get an idea of how the DBDs align (**Figure 3-1 B**), but we can visualize how the VirF DBD might interact with *pvirB* (**Figure 3-1 C**). The exact functions of each HTH motif in the AraC family of transcriptional regulators has been debated.[12,42,43] However, it seems clear that in addition to the amino acid side chain differences, slight variations in the orientations of both helices may be needed for each protein to specifically bind its cognate promoter.

Examination of the crystal structure of MarA in **Figure 3-3** reveals three amino acid side chains in helix-3 and four in helix-6 that interact with DNA. Each of these sidechains extend out of the helical core and make contacts with the *marRAB* promoter DNA bases directly or via an intervening water molecule (**Table 3-4**).[29,30] By aligning the primary VirF sequence and overlaying our homology models with the MarA structure, we were able to identify VirF DBD amino acids that correspond to these seven MarA residues (**Figure 3-3**).

Table 3-4: ***Specific Interactions Between MarA Residues and Bases Within the marRAB Promoter DNA Deduced from the Complex Structure (PBD:1BLO)****. Numbering corresponds to the numbering of the marRAB promoter sequence below, the bottom strand is in italics for clarity.*

^{01}GATTTAGCAAAACGTGGCAT^{20}

[40]CTAAATCGTTTTGCACCGTA[21]

MarA Residue	**W42**	**Q45**	**R46**	**Q92**	**T93**	**T95**	**R96**
VDW interactions	*C24, C25*, & T15	T15 & G14		T4 & T5		T4, T5 & H_2O-mediated to DNA backbone	*T32*
H-bond Interactions			G17*, G23 & C24*		*C34 & T35* both water-mediated		*G33*, G7 & A6 water-mediated

To probe the interactions of MarA with *marRAB*, we performed alanine scanning mutagenesis *in vitro,* on each of the seven DNA-interacting residues in MarA and evaluated their affinities for *marRAB* via electrophoretic mobility shift assays (EMSA). The K_D for MarA WT that we determined in this study diverges significantly from what was reported by Gillette *et al*.[30] It has been shown that deletions to the *marRAB* promoter, specifically to the binding site (i.e., Δ*marO281*) significantly reduce the affinity of MarA to DNA by up to 100-fold.[44] Hence, we tested binding of our purified WT MarA to both our 60 bp *marRAB* promoter and the 143 bp promoter used in Gillette et al.[30] (**Figure 3-9 A and B**). MarA affinities for both the 143 bp and 60 bp probes were not significantly different under our assay conditions and did not significantly vary when determined in the Gillette dialysis buffer (50 mM HEPES, 500 mM NaCl, 250 mM imidazole, 25% glycerol, pH 8.5, **Figure 3-9 C-F**). Upon running the EMSA according to the Gillette protocol, comparable binding affinities were again observed (**Figure 3-9 G and H**). In literature reports, differing experimental conditions and promoter lengths yielded K_D's varying from ~3 to 75 nM for the MarA•*marRAB* binding interaction.[30,44–48] While our K_D is higher than the highest previously reported K_D, it is consistent with our determination of the K_D for the DNA promoter used by Gillette et al.[30]

Overall, our observations are largely consistent with previously reported studies, in that alanine substitutions of R46 and R96 severely reduced MarA's affinity for *marRAB* DNA. While it is obvious that R46A decreased MarA's binding affinity (**Figure 3-6**), attempts to determine its K_D were unsuccessful (**Figure 3-7**). Interestingly, when substituting W42 to alanine, despite showing little decrease in MarA activity *in vivo*,[30] we observed a 2.3-fold decrease in binding affinity relative to WT (**Table 3-3**). The cumulative effect of losing the three VDW contacts that W42 makes with C31, C32, and T18 of the *marRAB* DNA are likely responsible for the decrease in affinity (**Table 3-4**). Substituting R96 for alanine resulted in a 4.1-fold decrease in binding affinity, the largest decrease of all the alanine substitution mutants (**Table 3-3**). In general, alanine substitutions made to helix-3 residues resulted in larger decreases in binding affinity than those in helix-6. While R96A dramatically decreased binding affinity, Q92A, T93A, and T95A each displayed an increase in binding affinity for *marRAB*. Q92A showed a surprising increase in affinity, increasing the K_D more than 20-fold compared to WT (**Table 3-3**). These findings align with observations made by Gillette *et al.* that the protein is more sensitive to alterations made to its N-terminal HTH (helices 2 and 3) than to alterations in the C-terminal HTH (helices 5 and 6).[30]

Overall, there were similar trends seen for both MarA and MalE-VirF mutants binding to their cognate promoters (**Table 3-3**), validating our MarA-based VirF DBD homology model. While both WT proteins bound with micromolar affinity to their cognate promoters, mutation of amino acids capable of hydrogen bonding or electrostatic interactions with their cognate promoter (MarA R46 and R96; MalE-VirF R192, K193, and R242) showed reductions or loss of DNA-binding activity. While R192 and K193 MalE-VirF alanine-mutants led to a complete loss of binding activity, R242A in MalE-VirF exhibited a >4-fold change in binding relative to WT

MalE-VirF but retained activity in both the FP and EMSA (**Table 3-3**). Similarly, R96A in MarA showed a 4-fold decrease in binding relative to MarA WT (**Figure 3-7**). Given that these residues are identical in both proteins and are likely to make electrostatic or hydrogen-bonding interactions with their DNA promoters, this similarity in affinity upon mutation is unsurprising.

In conclusion, we have generated models of the VirF DBD using GadX and MarA•*marRAB* structures. The two models of VirF, based on GadX and MarA, suggest a conformational change of the AraC proteins when they bind to DNA (**Figure 3-1**). If true, then screening for small molecule inhibitors of this conformational change might prevent DNA-binding and lead to a novel anti-virulence agent. Based on alignments with MarA, we tested seven alanine mutants within the DBD of the MalE-VirF fusion protein. Comparisons of the effects of the alanine mutants on DNA-binding validated the MarA-based VirF model and identified key interactions between VirF and one of its cognate promoters, *pvirB*. These DNA-binding assays, in conjunction with the observations we made of the VirF DBD through homology modeling, give a more detailed understanding of how VirF interacts with *pvirB*. Given the lack of structural data for this protein, these results and models will be useful in the continued efforts to analyze VirF's ability to bind its other promoters as well as identify and improve upon current DNA-binding inhibitors for the novel treatment of shigellosis. Moving forward, we can continue to study the MarA•*marRAB* interaction, further mutating the binding site to determine if it can be used as VirF•*virB* model.

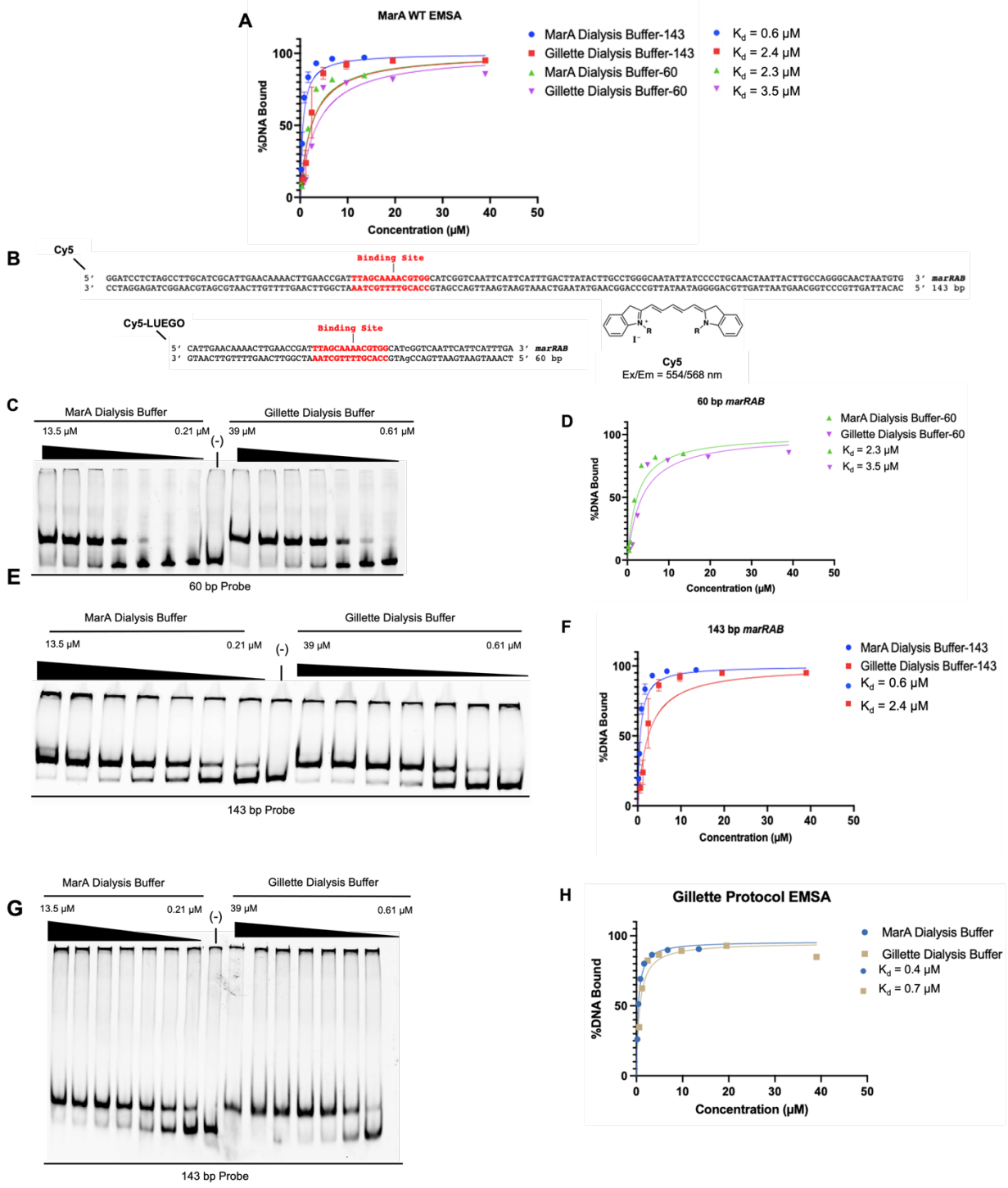


Figure 3-9: ***Validation of Experiments in Respect to the Literature.*** *A): Representative plot of MarA binding curves to the 60 bp or 143 bp marRAB probe in the MarA dialysis buffer from this work (50 mM Tris-HCl, 250 mM NaCl, 0.1% Triton X-100, 20% glycerol, pH 8.0) or the dialysis buffer used by Gillette et al. (50 mM HEPES, 500 mM NaCl, 250 mM imidazole, 25% glycerol, pH 8.5). KD values for the 143 bp probe were 0.5 μM and 2.0 μM for MarA Dialysis Buffer and Gillette Dialysis Buffer, respectively, and for the 60 bp probe the KD values were 2.0 μM* and 3.5 μM, respectively. B): 143 bp and 60 bp marRAB EMSA probes (See* ***Figure 3-5*** *for Cy5 and Cy5-LUEGO). Individual EMSA gels and binding curves for the 60 bp (C*

*and D) or the 143 bp (E and F). G and H): MarA EMSA with 143 bp probe ran using the reaction conditions of Gillette et al. (5% polyacrylamide gel, 1xTAE running buffer). KD values were 0.5 μM and 0.8 μM for MarA Dialysis Buffer and Gillette Dialysis Buffer with the 143 bp probe, respectively. (*For the 60 bp probe, we report a 2.0 μM KD in this set of experiments, whereas in Figure 4, we report a KD of 4.8 μM in a distinct set of experiments comparing WT and mutant MarA proteins.)*

Reference

(1) Stevens, P. Diseases of Poverty and the 10/90 Gap. In *Fighting the Diseases of Poverty*; 2004.

(2) Médecins Sans Frontières. *Fatal Imbalance: The Crisis in Research and Development for Drugs for Neglected Diseases*; 2001.

(3) Neill, J. O. '. Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations. Commissioned by the UK Prime Minister, 2014. **2014**, No. December.

(4) Murray, C. J.; Ikuta, K. S.; Sharara, F.; Swetschinski, L.; Robles Aguilar, G.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; et al. Global Burden of Bacterial Antimicrobial Resistance in 2019: A Systematic Analysis. *Lancet* **2022**, *399* (10325), 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0.

(5) Centers for Disease Control and Prevention. Antibiotic Resistance Threats in the United States. *U.S. Dep. Heal. Hum. Servies* **2019**, 1–113. https://doi.org/10.15620/cdc:82532.

(6) Williams, P.; Berkley, J. A. Dysentery (Shigellosis) Current WHO Guidelines and the WHO Essential Medicine List for Children http://www.who.int/selection_medicines/committees/expert/21/applications/s6_paed_antibiotics_appendix5_dysentery.pdf.

(7) Clatworthy, A. E.; Pierson, E.; Hung, D. T. Targeting Virulence: A New Paradigm for Antimicrobial Therapy. *Nat. Chem. Biol.* **2007**, *3* (9), 541–548. https://doi.org/10.1038/nchembio.2007.24.

(8) Nakayama, S. I.; Watanabe, H. Involvement of CpxA, a Sensor of a Two-Component Regulatory System, in the PH-Dependent Regulation of Expression of Shigella Sonnei

VirF Gene. *J. Bacteriol.* **1995**, *177* (17), 5062–5069. https://doi.org/10.1128/jb.177.17.5062-5069.1995.

(9) Porter, M. E.; Dorman, C. J. A Role for H-NS in the Thermo-Osmotic Regulation of Virulence Gene Expression in Shigella Flexneri. *J. Bacteriol.* **1994**, *176* (13), 4187–4191. https://doi.org/10.1128/jb.176.13.4187-4191.1994.

(10) Tobe, T.; Nagai, S.; Okada, N.; Adter, B.; Yoshikawa, M.; Sasakawa, C. Temperature-regulated Expression of Invasion Genes in Shigella Flexneri Is Controlled through the Transcriptional Activation of the VirB Gene on the Large Plasmid. *Mol. Microbiol.* **1991**, *5* (4), 887–893. https://doi.org/10.1111/j.1365-2958.1991.tb00762.x.

(11) Falconi, M.; Colonna, B.; Prosseda, G.; Micheli, G.; Gualerzi, C. O. 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.* **1998**, *17* (23), 7033–7043. https://doi.org/10.1093/emboj/17.23.7033.

(12) Gallegos, M. T.; Schleif, R.; Bairoch, A.; Hofmann, K.; Ramos, J. L. Arac/XylS Family of Transcriptional Regulators. *Microbiol. Mol. Biol. Rev.* **1997**, *61* (4), 393–410. https://doi.org/9409145.

(13) Martin, R. G.; Rosner, J. L. The AraC Transcriptional Activators. *Curr. Opin. Microbiol.* **2001**, No. 4, 132–137.

(14) Tobe, T.; Yoshikawa, M.; Mizuno, T.; Sasakawa, C. Transcriptional Control of the Invasion Regulatory Gene VirB of Shigella Flexneri: Activation by VirF and Repression by H-NS. *J. Bacteriol.* **1993**, *175* (19), 6142–6149.

(15) Adler, B.; Sasakawa, C.; Tobe, T.; Makino, S.; Komatsu, K.; Yoshikawa, M. A Dual

Transcriptional Activation System for the 230 Kb Plasmid Genes Coding for Virulence-Associated Antigens of Shigella Flexneri. *Mol. Microbiol.* **1989**, *3* (5), 627–635. https://doi.org/10.1111/j.1365-2958.1989.tb00210.x.

(16) Sakai, T.; Sasakawa, C.; Yoshikawa, M. Expression of Four Virulence Antigens of Shigella Flexneri Is Positively Regulated at the Transcriptional Level by the 30 Kilo Dalton VirF Protein. *Mol. Microbiol.* **1988**, *2* (5), 589–597. https://doi.org/10.1111/j.1365-2958.1988.tb00067.x.

(17) Bernardini, M. L.; Mounier, J.; D'Hauteville, H.; Coquis-Rondon, M.; Sansonetti, P. J. Identification of IcsA, a Plasmid Locus of Shigella Flexneri That Governs Bacterial Intra- and Intercellular Spread through Interaction with F-Actin. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86* (10), 3867–3871. https://doi.org/10.1073/pnas.86.10.3867.

(18) Tran, C. N.; Giangrossi, M.; Prosseda, G.; Brandi, A.; Di Martino, M. L.; Colonna, B.; Falconi, M. 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.* **2011**, *39* (18), 8122–8134. https://doi.org/10.1093/nar/gkr521.

(19) Giangrossi, M.; Prosseda, G.; Tran, C. N.; Brandi, A.; Colonna, B.; Falconi, M. A Novel Antisense RNA Regulates at Transcriptional Level the Virulence Gene IcsA of Shigella Flexneri. *Nucleic Acids Res.* **2010**, *38* (10), 3362–3375. https://doi.org/10.1093/nar/gkq025.

(20) Sansonetti, P. J.; Arondel, J.; Fontaine, A.; d'Hauteville, H.; Bernardini, M. L. OmpB (Osmo-Regulation) and IcsA (Cell-to-Cell Spread) Mutants of Shigella Flexneri: Vaccine Candidates and Probes to Study the Pathogenesis of Shigellosis. *Vaccine* **1991**, *9* (6), 416–422. https://doi.org/10.1016/0264-410X(91)90128-S.

(21) Theriot, J.; Portnoy, D. Theriot Lab at the University of Washington: Publications and Movies https://sites.uw.edu/theriotlab/movies/.

(22) Hurt, J. K.; McQuade, T. J.; Emanuele, A.; Larsen, M. J.; Garcia, G. A. High-Throughput Screening of the Virulence Regulator VirF: A Novel Antibacterial Target for Shigellosis. *J. Biomol. Screen.* **2010**, *15* (4), 379–387. https://doi.org/10.1177/1087057110362101.

(23) Emanuele, A. a; Adams, N. E.; Chen, Y.-C.; Maurelli, A. T.; Garcia, G. a. Potential Novel Antibiotics from HTS Targeting the Virulence-Regulating Transcription Factor, VirF, from Shigella Flexneri. *J. Antibiot. (Tokyo).* **2014**, *67* (5), 379–386. https://doi.org/10.1038/ja.2014.10.

(24) Emanuele, A. A.; Garcia, G. A. Mechanism of Action and Initial, In Vitro SAR of an Inhibitor of the Shigella Flexneri Virulence Regulator VirF. *PLoS One* **2015**, *10* (9), e0137410. https://doi.org/10.1371/journal.pone.0137410.

(25) Koppolu, V.; Osaka, I.; Skredenske, J. M.; Kettle, B.; Hefty, P. S.; Li, J.; Egan, S. M. Small-Molecule Inhibitor of the Shigella Flexneri Master Virulence Regulator VirF. *Infect. Immun.* **2013**, *81* (11), 4200–4207. https://doi.org/10.1128/IAI.00919-13.

(26) Jain, P.; Li, J.; Porubsky, P.; Neuenswander, B.; Egan, S. M.; Aubé, J.; Rogers, S. 3-Substituted Biquinolinium Inhibitors of AraC Family Transcriptional Activator VirF from S. Flexneri Obtained through in Situ Chemical Ionization of 3,4-Disubstituted Dihydroquinolines. *RSC Adv.* **2014**, *4* (75), 39809–39816. https://doi.org/10.1039/c4ra08384a.

(27) Schleif, R. AraC Protein: A Love-Hate Relationship. *BioEssays* **2003**, *25* (3), 274–282. https://doi.org/10.1002/bies.10237.

(28) Tramonti, A.; Visca, P.; De Canio, M.; Falconi, M.; De Biase, D. Functional

Characterization and Regulation of GadX, a Gene Encoding an AraC/XylS-like Transcriptional Activator of the Escherichia Coli Glutamic Acid Decarboxylase System? *J. Bacteriol.* **2002**, *184* (10), 2603–2613. https://doi.org/10.1128/JB.184.10.2603-2613.2002.

(29) Rhee, S.; Martin, R. G. R. G.; Rosner, J. L. J. L.; Davies, D. R. D. R. A Novel DNA-Binding Motif in MarA: The First Structure for an AraC Family Transcriptional Activator. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (18), 10413–10418. https://doi.org/10.1073/pnas.95.18.10413.

(30) Gillette, W. K.; Martin, R. G.; Rosner, J. L. Probing the Escherichia Coli Transcriptional Activator MarA Using Alanine-Scanning Mutagenesis: Residues Important for DNA Binding and Activation. *J. Mol. Biol.* **2000**, *299* (5), 1245–1255. https://doi.org/10.1006/jmbi.2000.3827.

(31) Rhee, S.; Martin, R. G.; Rosner, J. L.; Davies, D. R. A Novel DNA-Binding Motif in MarA: The First Structure for an AraC Family Transcriptional Activator. *Proc.Natl.Acad.Sci.USA* **1998**, *95*, 10413–10418. https://doi.org/10.2210/PDB1BL0/PDB.

(32) Guex, N.; Peitsch, M. C.; Schwede, T. Automated Comparative Protein Structure Modeling with SWISS-MODEL and Swiss-PdbViewer: A Historical Perspective. *Electrophoresis* **2009**, *30* (SUPPL. 1), 162–173. https://doi.org/10.1002/elps.200900140.

(33) Bertoni, M.; Kiefer, F.; Biasini, M.; Bordoli, L.; Schwede, T. Modeling Protein Quaternary Structure of Homo- and Hetero-Oligomers beyond Binary Interactions by Homology. *Sci. Rep.* **2017**, *7* (1), 1–15. https://doi.org/10.1038/s41598-017-09654-8.

(34) Bienert, S.; Waterhouse, A.; De Beer, T. A. P.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T. The SWISS-MODEL Repository-New Features and Functionality. *Nucleic*

Acids Res. **2017**, *45* (D1), D313–D319. https://doi.org/10.1093/nar/gkw1132.

(35) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; De Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res*. **2018**, *46* (W1), W296–W303. https://doi.org/10.1093/nar/gky427.

(36) Benkert, P.; Biasini, M.; Schwede, T. Toward the Estimation of the Absolute Quality of Individual Protein Structure Models. *Bioinformatics* **2011**, *27* (3), 343–350. https://doi.org/10.1093/bioinformatics/btq662.

(37) Jair, K. W.; Martin, R. G.; Rosner, J. L.; Fujita, N.; Ishihama, A.; Wolf, R. E. Purification and Regulatory Properties of MarA Protein, a Transcriptional Activator of Escherichia Coli Multiple Antibiotic and Superoxide Resistance Promoters. *J. Bacteriol*. **1995**, *177* (24), 7100–7104. https://doi.org/10.1128/jb.177.24.7100-7104.1995.

(38) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 Years of Image Analysis. *Nat. Methods* **2012**, *9* (7), 671–675. https://doi.org/10.1038/nmeth.2089.

(39) Software, G. P. Non-Linear Regression (One Site -- Specific Binding) Was Performed Using GraphPad Prism Version 9.0.0 for Mac OS X. San Diego, California USA.

(40) Software, G. P. Nonlinear Regression (Sigmoidal, 4PL, X Is Log(Concentration)) Was Performed Using GraphPad Prism Version 9.0.0 for Mac OS X. San Diego, California USA.

(41) Software GP. Normalize Was Performed Using GraphPad Prism Version 9.0.0 for Mac OS X. San Diego, California USA.

(42) Bhende, P. M.; Egan, S. M. *Amino Acid-DNA Contacts by RhaS: An AraC Family Transcription Activator*; 1999; Vol. 181.

(43) Dorman, C. J.; McKenna, S.; Beloin, C. Regulation of Virulence Gene Expression in Shigella Flexneri, a Facultative Intracellular Pathogen. *Int. J. Med. Microbiol.* **2001**, *291* (2), 89–96. https://doi.org/10.1078/1438-4221-00105.

(44) Martin, R. G.; Jair, K. W.; Wolf, R. E.; Rosner, J. L. Autoactivation of the MarRAB Multiple Antibiotic Resistance Operon by the MarA Transcriptional Activator in Escherichia Coli. *J. Bacteriol.* **1996**, *178* (8), 2216–2223. https://doi.org/10.1128/jb.178.8.2216-2223.1996.

(45) Kwon, H. J.; Bennik, M. H. J.; Demple, B.; Ellenberger, T. Crystal Structure of the Escherichia Coli Rob Transcription Factor in Complex with DNA. *Nat. Struct. Biol.* **2000**, *7* (5), 424–430. https://doi.org/10.1038/75213.

(46) Martin, R. G.; Rosner, J. L. Promoter Discrimination at Class I Mara Regulon Promoters Mediated by Glutamic Acid 89 of the Mara Transcriptional Activator of Escherichia Coli. *J. Bacteriol.* **2011**, *193* (2), 506–515. https://doi.org/10.1128/JB.00360-10.

(47) Martin, R. G.; Rosner, J. L. Binding of Purified Multiple Antibiotic-Resistance Repressor Protein (MarR) to Mar Operator Sequences. *Proc. Natl. Acad. Sci. USA* **1995**, *92* (12), 5456–5460.

(48) Martin, R. G.; Gillette, W. K.; Martin, N. I.; Rosner, J. L. Complex Formation between Activator and RNA Polymerase as the Basis for Transcriptional Activation by MarA and SoxS in Escherichia Coli. *Mol. Microbiol.* **2002**, *43* (2), 355–370. https://doi.org/10.1046/j.1365-2958.2002.02748.x.

Chapter 4 MarA-VirF Chimeric Proteins

Our lab has previously developed a *Shigella*-based β-galactosidase reporter assay in which VirF activates transcription of a reporter gene through the *virB* promoter.[1] This assay was used to conduct a high-throughput screen of 142,000 compounds from the University of Michigan Center for Chemical Genomics. Extensive screening triage allowed us to identify five promising compounds that could potentially inhibit VirF's activity in the β-galactosidase reporter assay as well as in bacterial cell invasion and plaque formation assays.[2] Of these compounds, we were able to elucidate the mechanism of action of one of them, 19615 (**Figure 3-8**). Through an electrophoretic mobility shift assay (EMSA), our lab determined that compound 19615 inhibits VirF by inhibiting DNA binding.[3] Attempts to generate SAR on these compounds were ultimately unsuccessful due to confounding issues with cell penetration and compound solubility. Additionally, the lack of structural information on VirF and the difficulties we report in obtaining active WT VirF in reasonable yields exacerbate any further attempts to study 19615. Due to this, we developed a homology model of the VirF DNA binding domain using MarA as a template. In **Chapter 3**, we detail our analysis of the VirF and the MarA DNA binding domains in an attempt to better understand how VirF recognizes and binds to the *virB* promoter.[4]

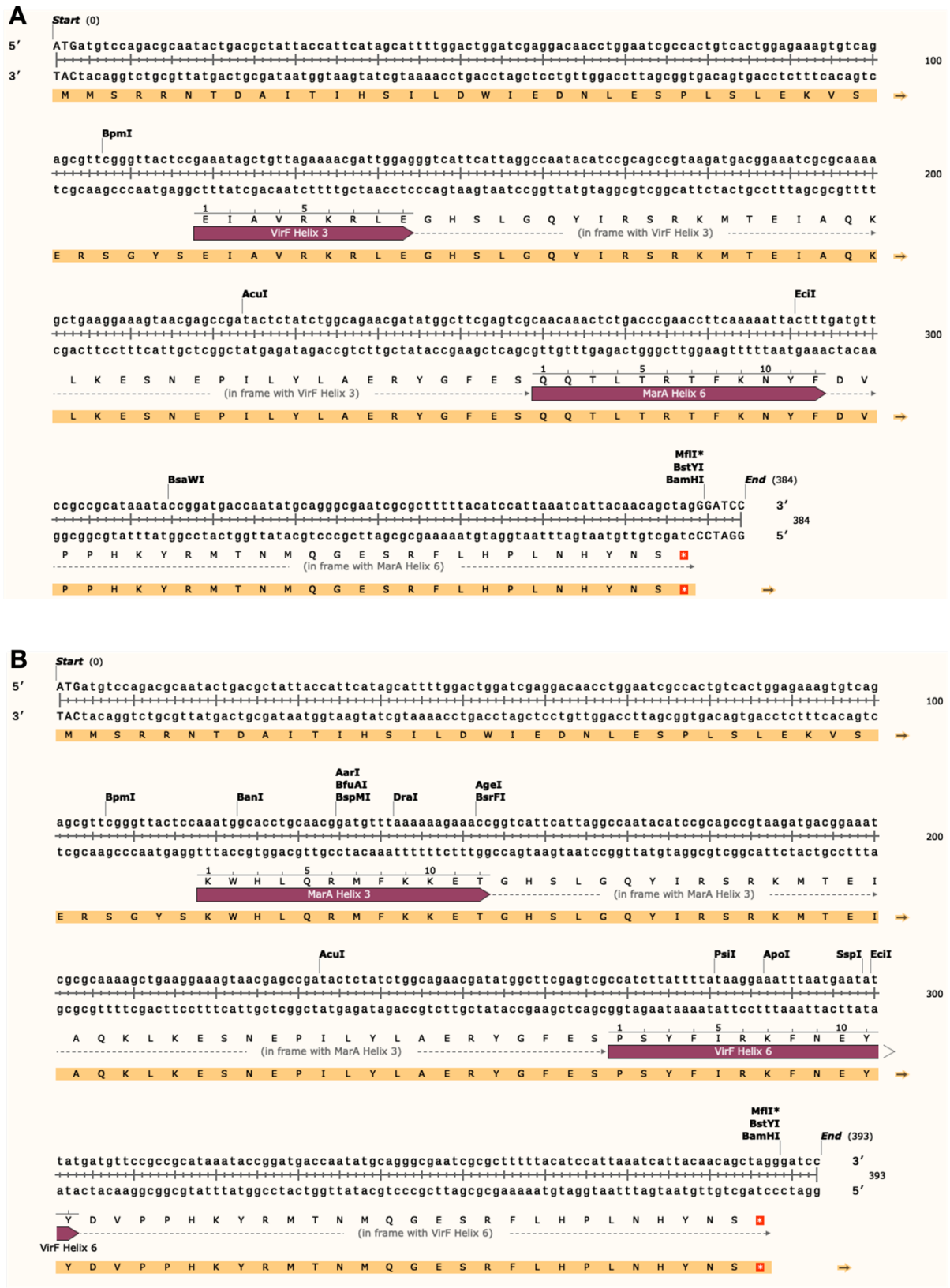
Here, we discuss using our homology model to design, express, and purify chimeric proteins that harbor the binding helices of VirF on a MarA peptide backbone in attempts to get MarA to recognize and bind to the *virB* promoter. This would allow us to use the chimeric

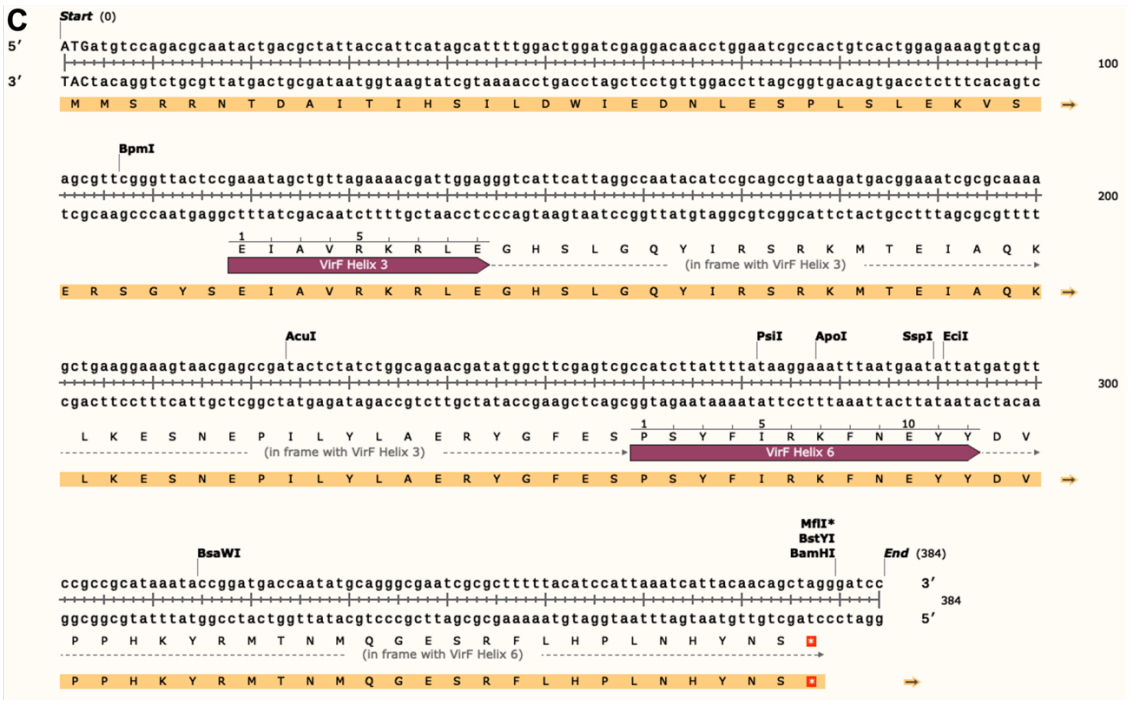
MarA as a model for *in vitro* analysis in lieu of a native WT VirF or truncated VirF DNA binding domain.

Materials and Methods

Reagents and Plasmids

All standard buffer components were purchased from Millipore Sigma or Thermo Fisher. Specific reagents or biological products not purchased from these are noted in parentheses. DNA oligonucleotides were purchased from Integrated DNA Technologies. The genes of the chimeric proteins were designed and submitted to Twist Bioscience for plasmid synthesis (**Figure 4-1**). The genes were inserted into the expression plasmid pET28a(+), chosen for its compatibility with isopropyl-β- D-thiogalactopyranoside (IPTG) induction and its N-terminal His-tag, and sent to us ready for expression. These vectors are referred to as *pET28aMarAVH3*, *pET28aMarAVH6*, *pET28aMarAVH3+6*, and *pET28aMarAV7*. Equipment utilized for these experiments was purchased from varying companies which are indicated in parentheses throughout this section.


A
Start (0)
5′ ATGatgtccagacgcaatactgacgctattaccattcatagcattttggactggatcgaggacaacctggaatcgccactgtcactggagaaagtgtcag
3′ TACtacaggtctgcgttatgactgcgataatggtaagtatcgtaaaacctgacctagctcctgttggaccttagcggtgacagtgacctctttcacagtc
100
M M S R R N T D A I T I H S I L D W I E D N L E S P L S L E K V S
BpmI
agcgttcgggttactccgaaatagctgttagaaaacgattggagggtcattcattaggccaatacatccgcagccgtaagatgacggaaatcgcgcaaaa
tcgcaagcccaatgaggctttatcgacaatcttttgctaacctcccagtaagtaatccggttatgtaggcgtcggcattctactgcctttagcgcgtttt
200
E I A V R K R L E G H S L G Q Y I R S R K M T E I A Q K
VirF Helix 3
(in frame with VirF Helix 3)
E R S G Y S E I A V R K R L E G H S L G Q Y I R S R K M T E I A Q K
AcuI
EciI
gctgaaggaaagtaacgagccgatactctatctggcagaacgatatggcttcgagtcgcaacaaactctgacccgaaccttcaaaaattactttgatgtt
cgacttcctttcattgctcggctatgagatagaccgtcttgctataccgaagctcagcgttgtttgagactgggcttggaagtttttaatgaaactacaa
300
L K E S N E P I L Y L A E R Y G F E S Q Q T L T R T F K N Y F D V
(in frame with VirF Helix 3)
MarA Helix 6
L K E S N E P I L Y L A E R Y G F E S Q Q T L T R T F K N Y F D V
MflI*
BstYI
BamHI
BsaWI
End (384)
ccgccgcataaataccggatgaccaatatgcagggcgaatcgcgcttttacatccattaaatcattacaacagctagGGATCC 3′
384
ggcggcgtatttatggcctactggttatacgtcccgcttagcgcgaaaaatgtaggtaatttagtaatgttgtcgatcCCTAGG 5′
P P H K Y R M T N M Q G E S R F L H P L N H Y N S
(in frame with MarA Helix 6)
P P H K Y R M T N M Q G E S R F L H P L N H Y N S
B
Start (0)
5′ ATGatgtccagacgcaatactgacgctattaccattcatagcattttggactggatcgaggacaacctggaatcgccactgtcactggagaaagtgtcag
3′ TACtacaggtctgcgttatgactgcgataatggtaagtatcgtaaaacctgacctagctcctgttggaccttagcggtgacagtgacctctttcacagtc
100
M M S R R N T D A I T I H S I L D W I E D N L E S P L S L E K V S
AarI
BfuAI
BspMI
BpmI
BanI
DraI
AgeI
BsrFI
agcgttcgggttactccaaatggcacctgcaacggatgtttaaaaaagaaaccggtcattcattaggccaatacatccgcagccgtaagatgacggaaat
tcgcaagcccaatgaggtttaccgtggacgttgcctacaaatttttttctttggccagtaagtaatccggttatgtaggcgtcggcattctactgccttta
200
K W H L Q R M F K K E T G H S L G Q Y I R S R K M T E I
MarA Helix 3
(in frame with MarA Helix 3)
E R S G Y S K W H L Q R M F K K E T G H S L G Q Y I R S R K M T E I
AcuI
PsiI
ApoI
SspI
EciI
cgcgcaaaagctgaaggaaagtaacgagccgatactctatctggcagaacgatatggcttcgagtcgccatcttattttataaggaaatttaatgaatat
gcgcgttttcgacttcctttcattgctcggctatgagatagaccgtcttgctataccgaagctcagcggtagaataaaatattcctttaaattacttata
300
A Q K L K E S N E P I L Y L A E R Y G F E S P S Y F I R K F N E Y
(in frame with MarA Helix 3)
VirF Helix 6
A Q K L K E S N E P I L Y L A E R Y G F E S P S Y F I R K F N E Y
MflI*
BstYI
BamHI
End (393)
tatgatgttccgccgcataaataccggatgaccaatatgcagggcgaatcgcgcttttacatccattaaatcattacaacagctagggatcc 3′
393
atactacaaggcggcgtatttatggcctactggttatacgtcccgcttagcgcgaaaaatgtaggtaatttagtaatgttgtcgatccctagg 5′
Y D V P P H K Y R M T N M Q G E S R F L H P L N H Y N S
VirF Helix 6
(in frame with VirF Helix 6)
Y D V P P H K Y R M T N M Q G E S R F L H P L N H Y N S


C
Start (0)
5′ ATGatgtccagacgcaatactgacgctattaccattcatagcattttggactggatcgaggacaacctggaatcgccactgtcactggagaaagtgtcag
3′ TACtacaggtctgcgttatgactgcgataatggtaagtatcgtaaaacctgacctagctcctgttggaccttagcggtgacagtgacctctttcacagtc
100
M M S R R N T D A I T I H S I L D W I E D N L E S P L S L E K V S
BpmI
agcgttcgggttactccgaaatagctgttagaaaacgattggagggtcattcattaggccaatacatccgcagccgtaagatgacggaaatcgcgcaaaa
tcgcaagcccaatgaggctttatcgacaatcttttgctaacctcccagtaagtaatccggttatgtaggcgtcggcattctactgcctttagcgcgtttt
200
1 5
E I A V R K R L E G H S L G Q Y I R S R K M T E I A Q K
VirF Helix 3
(in frame with VirF Helix 3)
E R S G Y S E I A V R K R L E G H S L G Q Y I R S R K M T E I A Q K
AcuI
PsiI
ApoI
SspI
EciI
gctgaaggaaagtaacgagccgatactctatctggcagaacgatatggcttcgagtcgccatcttatttttataaggaaatttaatgaatattatgatgtt
cgacttcctttcattgctcggctatgagatagaccgtcttgctataccgaagctcagcggtagaataaaatattcctttaaattacttataatactacaa
300
1 5 10
L K E S N E P I L Y L A E R Y G F E S P S Y F I R K F N E Y Y D V
(in frame with VirF Helix 3)
VirF Helix 6
L K E S N E P I L Y L A E R Y G F E S P S Y F I R K F N E Y Y D V
MflI*
BstYI
BamHI
BsaWI
End (384)
ccgccgcataaataccggatgaccaatatgcagggcgaatcgcgctttttacatccattaaatcattacaacagctagggatcc 3′
384
ggcggcgtatttatggcctactggttatacgtcccgcttagcgcgaaaaatgtaggtaatttagtaatgttgtcgatccctagg 5′
P P H K Y R M T N M Q G E S R F L H P L N H Y N S
(in frame with VirF Helix 6)
P P H K Y R M T N M Q G E S R F L H P L N H Y N S

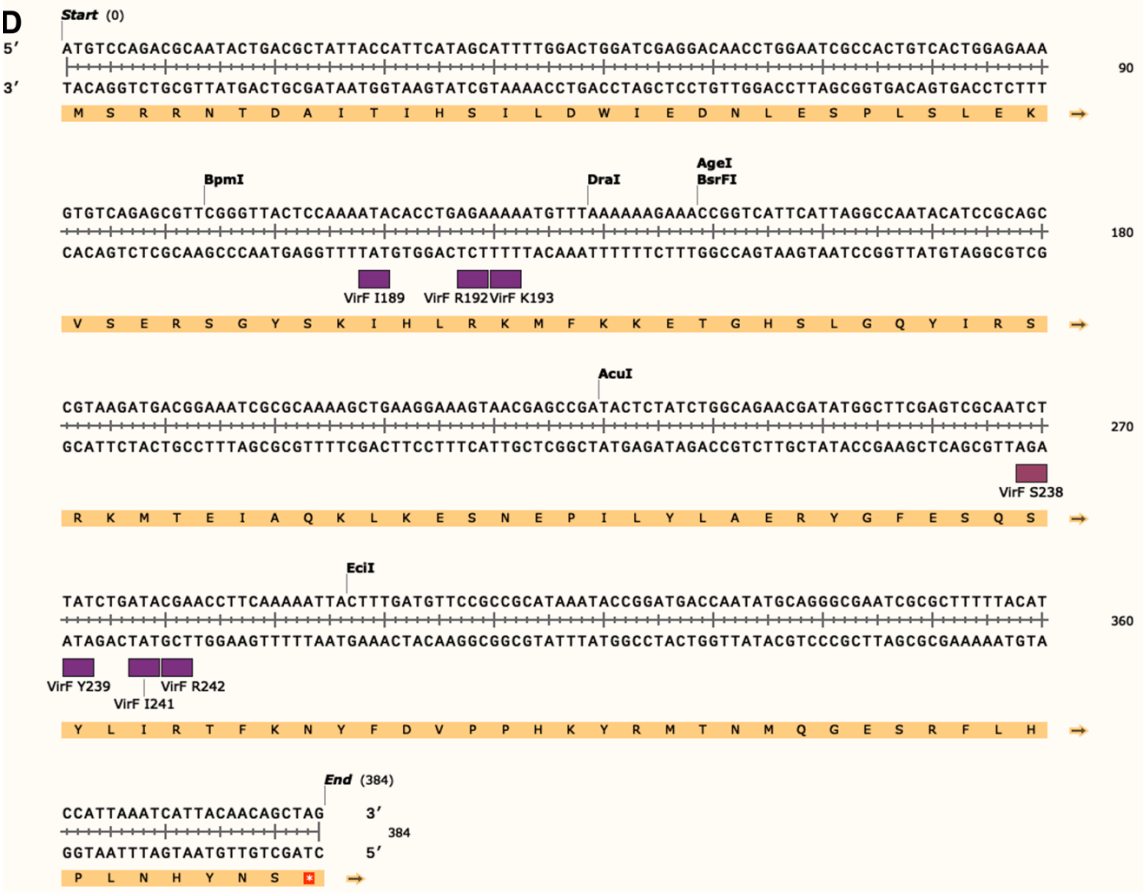

D
Start (0)
5′ ATGTCCAGACGCAATACTGACGCTATTACCATTCATAGCATTTTGGACTGGATCGAGGACAACCTGGAATCGCCACTGTCACTGGAGAAA
3′ TACAGGTCTGCGTTATGACTGCGATAATGGTAAGTATCGTAAAACCTGACCTAGCTCCTGTTGGACCTTAGCGGTGACAGTGACCTCTTT
90
M S R R N T D A I T I H S I L D W I E D N L E S P L S L E K
BpmI
DraI
AgeI
BsrFI
GTGTCAGAGCGTTCGGGTTACTCCAAAATACACCTGAGAAAAATGTTTAAAAAAGAAACCGGTCATTCATTAGGCCAATACATCCGCAGC
CACAGTCTCGCAAGCCCAATGAGGTTTTATGTGGACTCTTTTTACAAATTTTTTCTTTGGCCAGTAAGTAATCCGGTTATGTAGGCGTCG
180
VirF I189
VirF R192
VirF K193
V S E R S G Y S K I H L R K M F K K E T G H S L G Q Y I R S
AcuI
CGTAAGATGACGGAAATCGCGCAAAAGCTGAAGGAAAGTAACGAGCCGATACTCTATCTGGCAGAACGATATGGCTTCGAGTCGCAATCT
GCATTCTACTGCCTTTAGCGCGTTTTCGACTTCCTTTCATTGCTCGGCTATGAGATAGACCGTCTTGCTATACCGAAGCTCAGCGTTAGA
270
VirF S238
R K M T E I A Q K L K E S N E P I L Y L A E R Y G F E S Q S
EciI
TATCTGATACGAACCTTCAAAAATTACTTTGATGTTCCGCCGCATAAATACCGGATGACCAATATGCAGGGCGAATCGCGCTTTTTACAT
ATAGACTATGCTTGGAAGTTTTTAATGAAACTACAAGGCGGCGTATTTATGGCCTACTGGTTATACGTCCCGCTTAGCGCGAAAAATGTA
360
VirF Y239
VirF R242
VirF I241
Y L I R T F K N Y F D V P P H K Y R M T N M Q G E S R F L H
End (384)
CCATTAAATCATTACAACAGCTAG 3′
384
GGTAATTTAGTAATGTTGTCGATC 5′
P L N H Y N S

Figure 4-1: ***Chimeric MarA-VirF Gene Sequences.*** *A): MarA with helix 3 substituted for helix 3 of VirF. B): MarA with helix 6 substituted for helix 6 of VirF..C): MarA with helix 3 and 6 substituted with helix 3 and 6 of VirF. D): MarA with select mutations (shown in purple) to match the amino acids of VirF studied in* ***Chapter 3****.*

Expression Testing

Expression tests were performed on the three expression vectors *pET28aMarAVH3*, *pET28aMarAVH6*, and *pET28aMarAVH3+6*. First, 5 mL cultures of 2xTY media (16 g bactotryptone, 10 g yeast extract, 5 g NaCl per liter of water supplemented with 100 μg/mL carbenicillin) containing one of the three expression vectors were grown overnight and used to inoculate a fresh culture the following day. The cultures were incubated at 37 °C with agitation until the OD_{600} reached 0.8. The cultures were induced with IPTG at a final concentration of 0.4 mM and continued to grow overnight under the same conditions. The following day, the cultures were pelleted, and the cells were resuspended in 25 mL of lysis buffer (50 mM Tris-HCl, 1 mM EDTA, 1 M NaCl, pH 7.5) supplemented with supplemented with a "c0mplete, Mini Protease Inhibitor Cocktail" tablet (Roche) and 0.1 mM phenylmethylsulphonyl fluoride (PMSF). The mixture was sonicated and centrifuged again before the soluble and insoluble portions were evaluated by SDS PAGE.

Chimeric Protein Purification

Purification of the MarA-VirF chimeric proteins was performed similarly to the purification of MarA WT and the DNA binding domain (DBD) mutants as described in **Chapter 3**.[5] Starter cultures (10 mL) of *E. coli* BL21(DE3) containing one of the *pEt28a* expression plasmids was grown overnight in 2xTY media broth supplemented with carbenicillin at 37 °C under vigorous agitation. The next day, the starter culture was used to inoculate 1 L of 2xTY broth supplemented with carbenicillin. The cells were grown to an $OD_{600} = 0.8$ before

expression was induced with the addition of IPTG at a final concentration of 0.4 mM and the culture continued to shake overnight at 16 °C. Cells were then harvested by centrifugation (6000 x g, 4 °C, 15 min) before being resuspended in 25 mL lysis buffer (as described above). All proceeding steps were performed on ice or at 4 °C. Cells were lysed via sonication (8 cycles, 15 sec pulse, 3 min intervals, 60 % setting) utilizing an ultrasonic XL2020 sonicator (Misonix). Following sonication, the solution was pelleted by ultracentrifugation (120,000 x g, 4 °C, 30 min). The supernatant was discarded, and the pellet was washed with 30 mL of denature buffer 1 (50 mM Tris-HCl, 4 M urea, pH 8.5) before repeating the ultracentrifugation. The supernatant was discarded again, and the pellet was resuspended in 25 mL denature buffer 2 (50 mM Tris-HCl, 6 M guanidinium chloride, pH 8.5). The mixture was subjected to ultracentrifugation a third time, collecting the supernatant. Next, 2 mL of Ni-NTA agarose (Qiagen) were added, and the solution was rocked gently overnight at 4 °C. The following day, the resin slurry was poured into an empty purification column. Once settled, the resin bed was washed stepwise with increasing concentrations of imidazole by combining elution buffer (50 mM Tris-HCl, 500 mM NaCl, 1 M imidazole, pH 8.5) and wash buffer (50 mM Tris-HCl, 500 mM NaCl, pH 8.5) to 100 mM, 300 mM, 500 mM, and 750 mM imidazole. The fraction eluted with 300 mM imidazole contained the majority of the chimeric protein of interest as determined by SDS PAGE. The protein was dialyzed (50 mM Tris-HCl, 250 mM NaCl, 0.1% Triton X-100, 20% glycerol, pH 8.0) in two overnight steps to remove the imidazole.

Electrophoretic Mobility Shift Assay (EMSA)

EMSAs were performed as previously described in Emanuele and Garcia, 2015.[3] Prior to preparing the gel and reactions, the Cy5-labeled *virB* (*pvirB*) and *marRAB* oligonucleotide

promoter probes were annealed as previously described using the oligonucleotide primers found in **Table 3-2.** A 6 % native polyacrylamide gel was prepared using 30 % acrylamide/bis-acrylamide (29:1 ratio) solution and TBE buffer (0.25x final concentration; 22 mM Tris Base, 22 mM boric acid, 0.5 mM EDTA, pH 8.5 or 9.5). Prior to running the reactions, the empty gel was electrophoresed for 1 hour at 150 V in 0.25x TBE buffer (pH 8.5). Each reaction was composed of 3 µL of probe (42 nM, *marRAB* probe or *pvirB* probe), 9 µL of chimeric protein (MarA VH3 – 8.75 µM, MarA VH6 – 12.5 µM, MarA VH3+6 – 28.5 µM, MarA V7 – 8-0.25 µM), 1 µL salmon sperm DNA (Invitrogen; 0.7 mg/mL), 0.5 µL BSA (0.07 mg/mL), and 1.5 µL Milli-Q H_2O for a final volume of 15 µL. Final concentrations of DNA probe, salmon sperm DNA, and BSA within each reaction are presented in parentheses. Final buffer conditions were 40 mM Tris-HCl, 160 mM NaCl, 0.04 mM EDTA, 0.06 % Triton X-100, 12 % glycerol, pH 8.0.

Results

Helix-Turn-Helix Swapping

The chimeric proteins used in these experiments were designed by evaluating the superposition of our VirF DBD homology model and the MarA•*marRAB* crystal structure (PDB: 1BL0).[6] As seen in **Figure 3-1** and **Figure 3-3**, the helix-turn-helix motifs at each of the two binding sites overlay well. This led us to hypothesize that it might be possible to swap the helices that insert into the major grooves of DNA when binding. We designed three chimeric MarA proteins incorporating different combinations of the VirF DBD helix-turn-helix binding site. These were MarA with VirF helix 3 (MarA VH3), MarA with VirF helix 6 (MarA VH6), and MarA with both VirF Helix 3 and 6 (MarA VH3+6) (**Figure 4-1**).

Once the expression vectors were received from Twist Bioscience, we tested the expression of these proteins under similar conditions to MarA WT expression and purification. As seen previously with MarA WT and the DBD mutants studied in **Chapter 3**, the chimeric proteins were in the insoluble lysate (**Figure 4-2**). Additionally, MarA VH3+6 showed noticeably lower expression relative to the other two chimeric proteins. Still, we decided to move forward and purify each of the proteins.

The same protocol that was used to obtain pure and active MarA in **Chapter 3** was utilized to express and purify the chimeric MarA-VirF proteins from the cell pellet.[5] Due to the guanidinium HCl and urea present throughout the purification, SDS PAGE analysis of each step could not be visualized. Instead, we evaluated the size of our eluted fractions only. In **Figure 4-3 A**, a band can be seen quite clearly for MarA VH3 and MarA VH3+6 but not for MarA VH6. The purification of MarA VH6 was repeated, and the gel is shown in **Figure 4-3 B**.

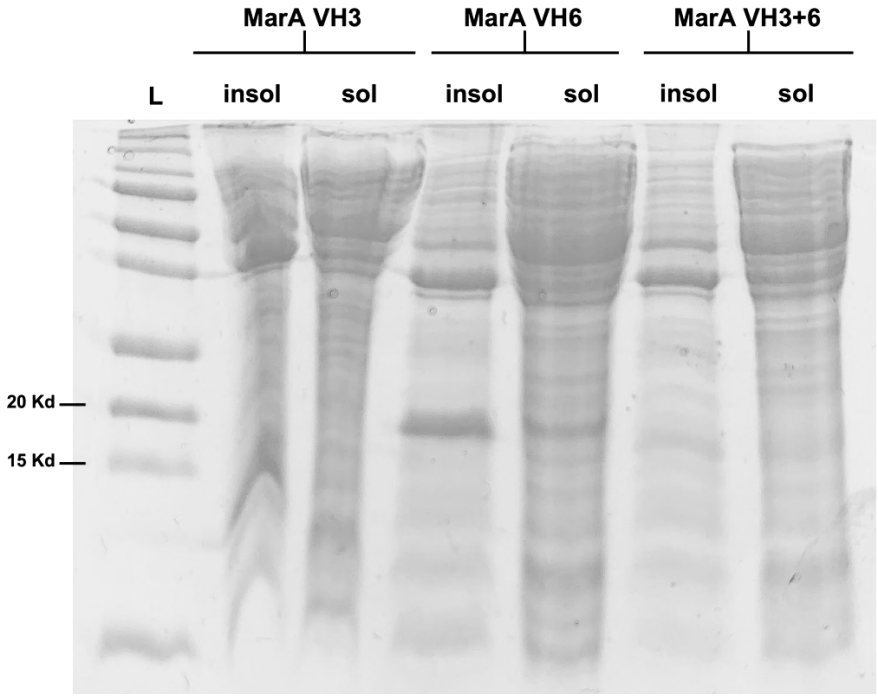


Figure 4-2: ***Expression Gel for Three MarA-VirF Chimeric Proteins.*** *The proteins must be purified from the insoluble pellet.*

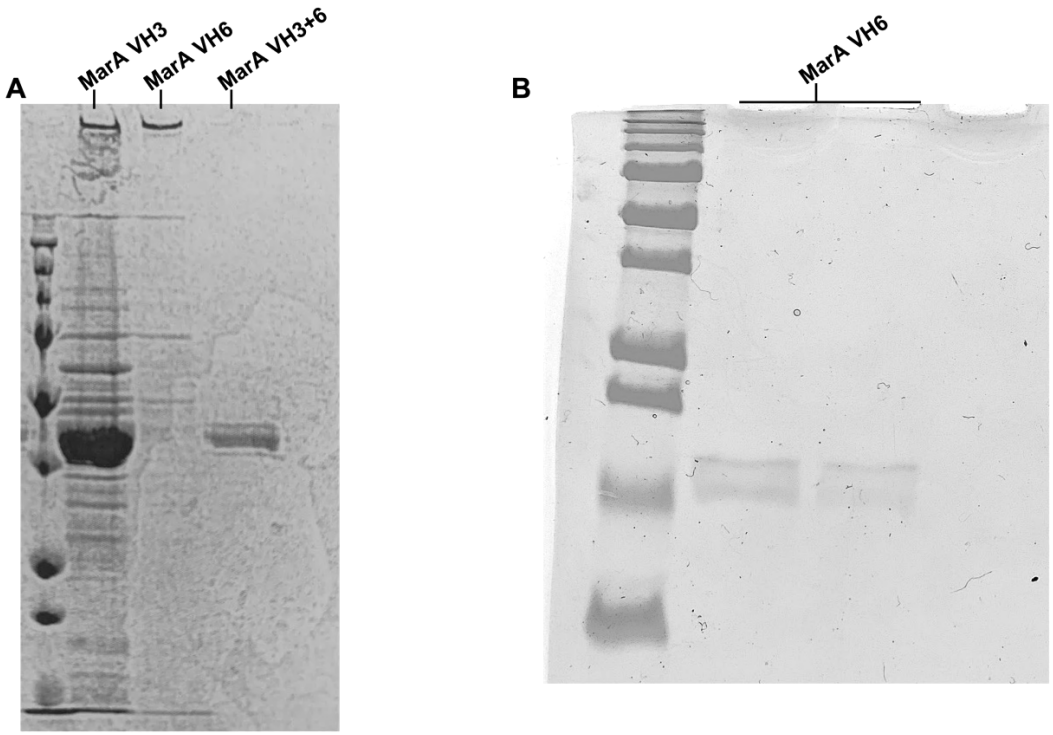


Figure 4-3: ***Purification Gels for Three MarA-VirF Chimeric Proteins.*** *A): Gel of three chimeric proteins. MarA VH6 was repeated and is shown in B).*

The activity of these chimeric proteins was evaluated by EMSA, using a modified protocol from what was described in **Chapter 3**. The three helix swapped proteins were evaluated for their capability to recognize and bind to the cognate promoters of MarA and VirF, *marRAB and virB*, respectively. Unfortunately, no activity was evident for either promoter by the helix swapped proteins (**Figure 4-4**).

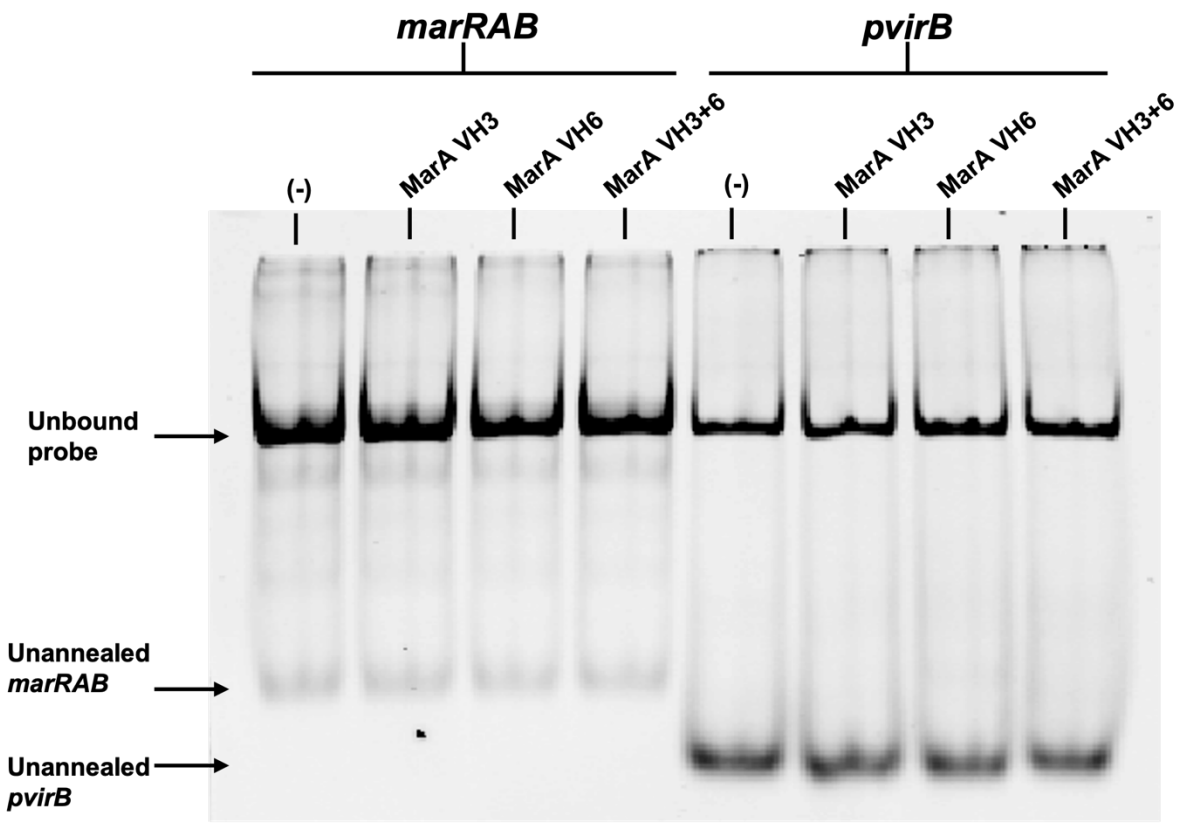


Figure 4-4: ***EMSA to Evaluate Helix-Swapped Chimeric Proteins' Activity.*** *None of the mutants gained affinity for pvirB and all lost affinity for marRAB.*

MarA-VirF Binding Site Mutant

Along with our helix-turn-helix swapped mutants, we designed a mutant of MarA that contained six mutations in the DNA binding domain that correspond to the VirF residues that were evaluated in **Chapter 3** and in Ragazzone et al.[4] Six mutations were made (**Figure 4-1 D**) to MarA: W42I, Q45R, R46K, Q92S, T93Y, and T95I (the 7th residue was arginine for both proteins). This mutant, referred to as MarA V7, was designed and ordered from Twist Bioscience in a pET28a expression vector. The expression and purification were performed following the same protocol as the other three chimeric proteins described here. MarA V7 was evaluated in an EMSA for its capability to recognize and bind to *marRAB* and *pvirB*. Similar to

the individual helix swapped mutants, MarA V7 exhibited no observable shift in the EMSA that would indicate binding to either of the two promoters (**Figure 4-5**).

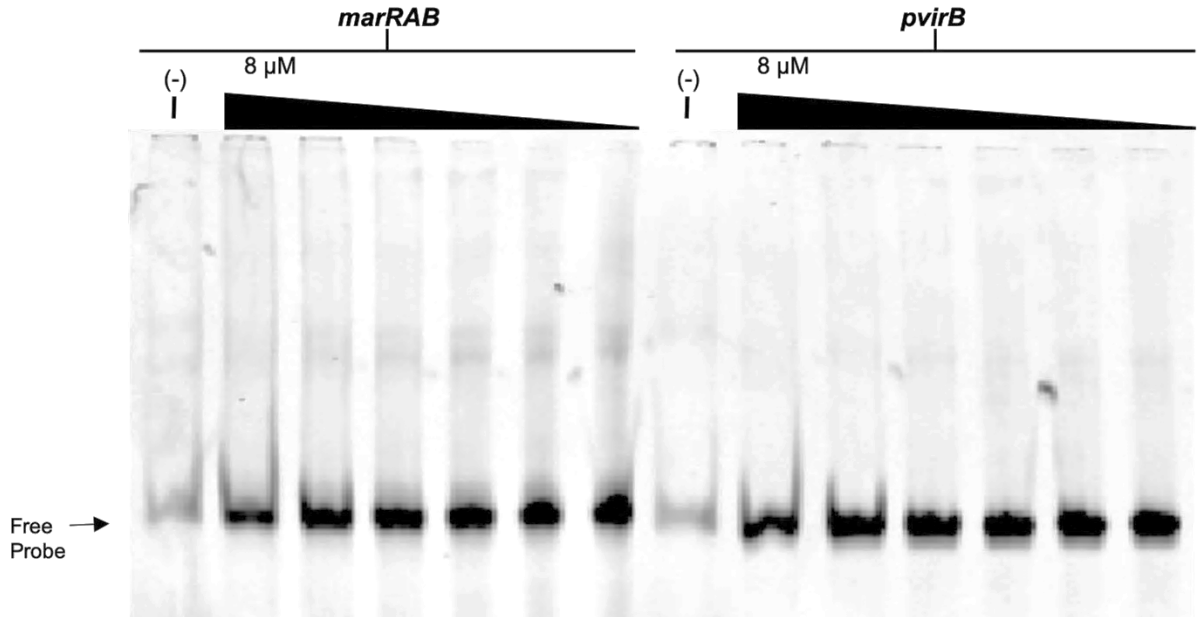


Figure 4-5: ***EMSA to Evaluate the DNA Binding Activity of MarA V7.***

Discussion

The use of protein engineering to create hybrid or chimeric proteins has been used in a few cases in the literature to study protein pathways and activity.[7–9] However, not many examples exist that use this technique to alter the specificity of DNA sequence recognition. In Ebright et al.[10], they describe making mutations to the catabolite gene activator protein (CAP) that resulted in a change of specificity for A•T base pairs rather than G•C at the same location of the DNA sequence. Still, this example is only altering the recognition of two bases in specific locations of the DNA sequence. We wanted to go beyond this and create a mutant protein that would recognize an alternative sequence. Specifically, we wanted to create a mutant version of MarA that would have the capability to bind the *virB* promoter. This would be immensely

beneficial in studying inhibitors of the VirF•*virB* interaction, such as our compound 19615 (**Figure 3-8 A**), considering the challenges in isolating active WT VirF for *in vitro* analysis. The rational for this largely lies in the similarity of the DNA binding domains of MarA and VirF, as well as similarities to most of the AraC family members.[11,12]

Although the CDs of our chimeric proteins were not perfect, we believe there were enough helical features to show that each of the chimeric MarA were successfully expressed and purified (**Figure 4-6**). We were not able to get a spectrum for MarA V7. Unfortunately, none of these individual helix-swapped mutants exhibited any observable activity for either of the two promoter sequences. Changing one of the MarA helices that insert into the major groove of DNA during binding was enough to lose all affinity for *marRAB* (**Figure 4-4**). Additionally, swapping both helices for the VirF amino acid sequences while keeping the MarA backbone was not sufficient in *virB* promoter recognition. It is possible that *virB* sequence recognition can be achieved eventually by creating a library of MarA mutants. It is possible that the side chains protruding from the binding helices in MarA VH3, MarA VH6, and MarA VH3+6 are in the wrong orientation to make the correct contacts with DNA, and by "ratcheting" the helices one residue at time, eventually the right orientation may be achieved. However, this would be a challenging project to take on, likely needing dozens of variations to the MarA-VirF binding helices.

MarA V7 harbored six point mutations in the MarA binding site to match the amino acids in VirF that were thought to be making base contacts when binding *pvirB* (**Figure 4-1 D**). This was different from the previous chimeric proteins tested in that most of helices 3 and 6 still corresponded to the MarA sequence. When evaluating the binding, we saw that changing the specific residues that make contact with the DNA was not enough to make MarA recognize and

bind to *pvirB* (**Figure 4-5**). If we manipulate the brightness and contrast of the gel, it's possible one can convince themselves that there is an extremely faint band in the lanes corresponding to marRAB. However, the trend of the band is counter intuitive of the protein concentration (the "band" would be growing darker with decreasing protein concentrations), and if time permitted it may be worth investigating this more. With more time and resources, it would have been interesting to see how slowly incorporating these mutations one at a time would have affected binding. It's possible that the orientations of these point mutations have a similar issue to the full helix swapped mutants and are not oriented in space correctly to make contact. It would be interesting to continue to probe MarA with mutations to see if binding to the *virB* promoter can be achieved.

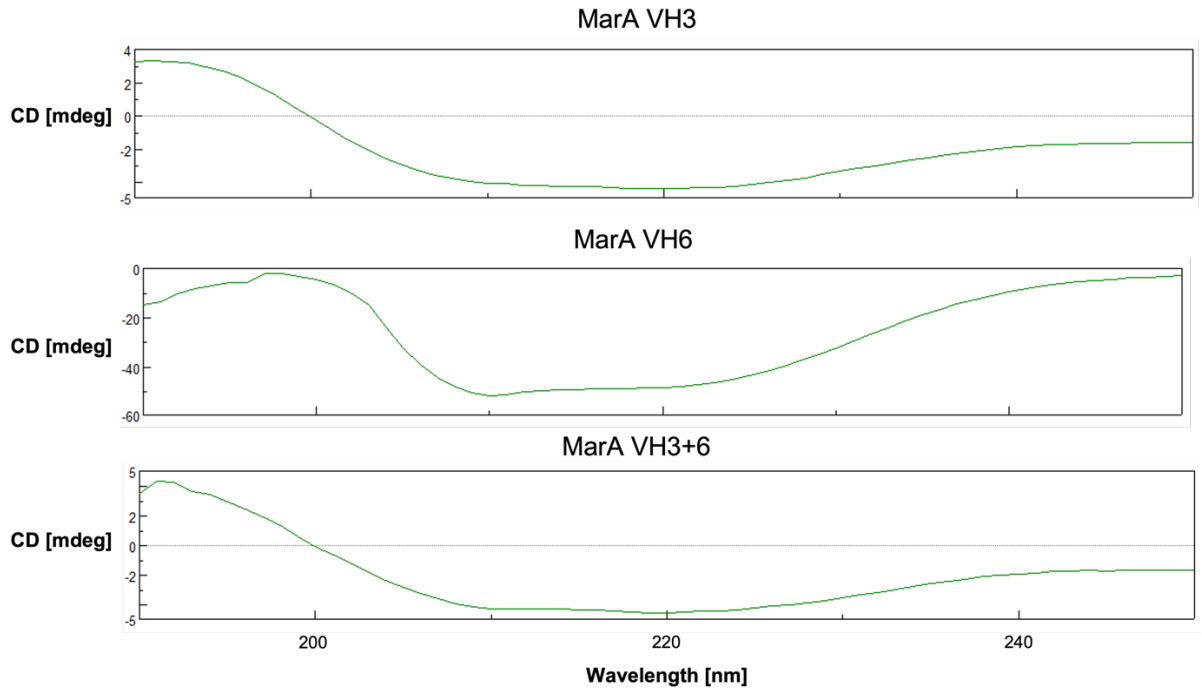


Figure 4-6: ***Circular Dichroism Spectra of Each of the Full Helix Swapped Chimeric Proteins.***

References

(1) Hurt, J. K.; McQuade, T. J.; Emanuele, A.; Larsen, M. J.; Garcia, G. A. High-Throughput Screening of the Virulence Regulator VirF: A Novel Antibacterial Target for Shigellosis. *J. Biomol. Screen.* **2010**, *15* (4), 379–387. https://doi.org/10.1177/1087057110362101.

(2) Emanuele, A. a; Adams, N. E.; Chen, Y.-C.; Maurelli, A. T.; Garcia, G. a. Potential Novel Antibiotics from HTS Targeting the Virulence-Regulating Transcription Factor, VirF, from Shigella Flexneri. *J. Antibiot. (Tokyo).* **2014**, *67* (5), 379–386. https://doi.org/10.1038/ja.2014.10.

(3) Emanuele, A. A.; Garcia, G. A. Mechanism of Action and Initial, In Vitro SAR of an Inhibitor of the Shigella Flexneri Virulence Regulator VirF. *PLoS One* **2015**, *10* (9), e0137410. https://doi.org/10.1371/journal.pone.0137410.

(4) Ragazzone, N. J.; Dow, G. T.; Garcia, G. A. Elucidation of Key Interactions between VirF and the VirB Promoter in Shigella Flexneri Using E. Coli MarA- and GadX-Based Homology Models and In Vitro Analysis of the DNA-Binding Domains of VirF and MarA. *J. Bacteriol.* **2022**, *204* (9), e0014322. https://doi.org/10.1128/jb.00143-22.

(5) Jair, K. W.; Martin, R. G.; Rosner, J. L.; Fujita, N.; Ishihama, A.; Wolf, R. E. Purification and Regulatory Properties of MarA Protein, a Transcriptional Activator of Escherichia Coli Multiple Antibiotic and Superoxide Resistance Promoters. *J. Bacteriol.* **1995**, *177* (24), 7100–7104. https://doi.org/10.1128/jb.177.24.7100-7104.1995.

(6) Rhee, S.; Martin, R. G. R. G.; Rosner, J. L. J. L.; Davies, D. R. D. R. A Novel DNA-Binding Motif in MarA: The First Structure for an AraC Family Transcriptional Activator. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (18), 10413–10418. https://doi.org/10.1073/pnas.95.18.10413.

(7) Ostermeier, M.; Nixon, A. E.; Shim, J. H.; Benkovic, S. J. Combinatorial Protein Engineering by Incremental Truncation. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (7), 3562–3567. https://doi.org/10.1073/pnas.96.7.3562.

(8) O'Maille, P. E.; Bakhtina, M.; Tsai, M. D. Structure-Based Combinatorial Protein Engineering (SCOPE). *J. Mol. Biol.* **2002**, *321* (4), 677–691. https://doi.org/10.1016/S0022-2836(02)00675-7.

(9) Rodgers, M. E.; Schleif, R. Solution Structure of the DNA Binding Domain of AraC Protein. *Proteins Struct. Funct. Bioinforma.* **2009**, *77* (1), 202–208. https://doi.org/10.1002/prot.22431.

(10) Ebright, R. H.; Cossart, P.; Gicquel-Sanzey, B.; Beckwith, J. Molecular Basis of DNA Sequence Recognition by the Catabolite Gene Activator Protein: Detailed Inferences from Three Mutations That Alter DNA Sequence Specificity. *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81* (23 I), 7274–7278. https://doi.org/10.1073/pnas.81.23.7274.

(11) Gallegos, M. T.; Schleif, R.; Bairoch, A.; Hofmann, K.; Ramos, J. L. Arac/XylS Family of Transcriptional Regulators. *Microbiol. Mol. Biol. Rev.* **1997**, *61* (4), 393–410. https://doi.org/9409145.

(12) Jonathan Posner, James A. Russell, and B. S. P. MarA, SoxS and Rob of Escherichia Coli – Global Regulators of Multidrug Resistance, Virulence and Stress Response. *Int J Biotechnol Wellness Ind* **2008**, *2* (3), 101–124. https://doi.org/10.6000/1927-3037.2013.02.03.2.MarA

Chapter 5 Caco-2 Invasion and Plaque Formation Assays

In collaboration with GlaxoSmithKline (GSK), our lab developed an assay for the inhibition of intra-macrophage survival of *Shigella flexneri*. This assay was used to conduct a high throughput screen of 1.7 million compounds to identify antivirulence compounds that have the capability to prevent *Shigella* from inducing apoptosis of the macrophages. These antivirulence hits would potentially avoid the selective pressure that leads to the rise of antibiotic resistance. The screen led to the identification of 23,118 initial hits. After reconfirmation screens and dose response assays, as well as a secondary screen evaluating toxicity, the number of hits was refined to 44.

Here, we set out to reproduce Caco-2 cell based assays[2] that would verify that these hits are virulence inhibitors. First would be to conduct an invasion assay that would evaluate the capability of these hit compounds to inhibit the bacteria's ability to invade the Caco-2 monolayers. Additionally, a plaque formation assay would be conducted that would give insight to whether these hits were targeting *Shigella*'s ability to spread from cell to cell.

Materials and Methods

Reagents and Cell lines

All reagents were purchased from Sigma-Aldrich unless otherwise specified. Tryptic soy broth, carbenicillin and microtiter plates used in these experiments were purchased from Fisher Scientific. Wild-type *Shigella flexneri* serotype 2a strain, 2457T was purchased from ATCC.

The isogenic virulence plasmid-cured derivative, BS103, was a gift to our lab from Dr. Anthony Maurelli.[1] All tissue culture media including Dulbecco's modified Eagle medium (DMEM), phosphate buffered saline (PBS), Roswell Park Memorial Institute Medium (RPMI), and fetal bovine serum (FBS) was purchased from Life Technologies. Caco-2 cells were received as a gift from Prof. J. Sexton and also were later ordered from ATCC.

Shigella Gentamicin Protection Invasion Assay

The *Shigella* invasion assays were performed as previously described.[2] Caco-2 cells were seeded in six-well plates 2-3 days prior to the start of the experiments to ensure cells are grown to near confluency in time. *Shigella flexneri* 2457T or avirulent *Shigella flexneri* BS103 were grown overnight, subcultured into TSB at a 1:100 dilution, and grown at 37 °C with agitation for 2.5 hours. Strains were standardized to an OD_{600} of 0.35, washed in PBS and resuspended in DMEM. The input bacteria were titered on TSB Congo Red plates before applying to six-well plates seeded to semiconfluency with Caco-2 cells. The plates were incubated at 37 °C under 5 % CO_2 for 2 hours to allow invasion to take place. The monolayers were then gently washed 4x with DPBS before 2 mL of DMEM supplemented with gentamicin (50 μg/mL) to each well. The plates were incubated for another 30 minutes under the same conditions. The monolayers were washed again with DPBS before an additional 2 mL of DMEM supplemented with gentamicin (50 μg/mL) and incubated for another hour. The monolayers were washed four more times with DPBS, and 1 mL of 0.5 % triton X-100 was added to each well. The plates were incubated for 5 minutes at 37 °C before the monolayers were scraped with a pipette tip and collected in individual Eppendorf tubes. The solutions were

vortexed for 30 seconds and titered in 1 mL buffered saline gelatin (BSG) cultures. From each dilution, 100 µL were plated onto Congo Red plates to determine output colony forming units.

$$Invasion\ \% = \frac{total\ recovered\ bacteria}{total\ input\ bacteria} x100$$

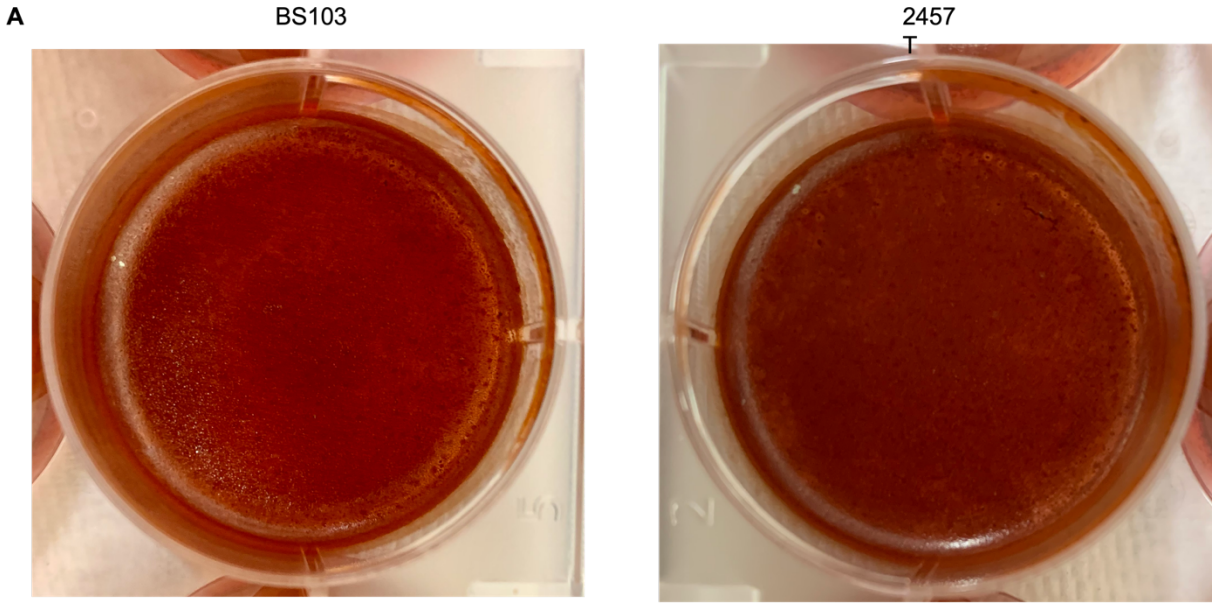
Plaque Formation Assay

The *Shigella* plaque formation assays were performed as previously described.[2] In brief, Caco-2 cells were seeded 3-4 days prior to the start of the assay to ensure cells are grown to confluence at the time of the experiment. *Shigella flexneri* 2457T or avirulent *Shigella flexneri* BS103 were grown overnight, subcultured into Tryptic Soy Broth (TSB) at a 1:100 dilution, and grown at 37 °C with agitation for 2.5 hours. Strains were standardized to an OD_{600} of 0.35, washed in PBS and diluted in DMEM. The input bacteria were titered on TSB Congo Red plates before applying to plates seeded to confluence with Caco-2 cells. The plates were gently rocked to mix before they were incubated at 37 °C under 5 % CO_2 for 2 hours. An agarose solution consisting of DMEM, 10 % fetal bovine serum, 50 µg/mL gentamicin, 0.5 % agarose was then applied to the monolayers. The plates were incubated at 37 °C under 5 % CO_2 for 3 days and then stained with 0.5 % Neutral Red to visualize the plaques. Efficiency of plaque formation was calculated by dividing the total number of plaques observed post-staining by the total number of input bacteria.

$$Plaque\ Efficiency\ (\%) = \frac{total\ number\ of\ plaques}{total\ input\ bacteria} x100$$

Results

Reproducing the *Shigella* invasion and plaque formation assays proved to be more challenging than originally anticipated. We received a flask of Caco-2 cells from Dr. Sexton's lab to start culturing in our BSL-2 lab to use in these experiments. Our first few trials of the plaque formation assay were performed using a smaller volume six-well plate. We also used our cryo-stock of *S. flexneri* 2457T for invasion in the early attempts at this experiment. After incubating the monolayers with either 2457T or BS103 for three days, no plaques were observed in the plates (**Figure 5-1 A**). Analyzing the plates of input bacteria, we thought it was possible that the *S. flexneri* 2457T had lost its virulence plasmid. *Shigella* colonies plated on Congo Red plates should have a red hue if the bacteria is virulent and a white hue if the bacteria is avirulent.[1] This distinction was unclear in our input colonies (**Figure 5-1 B**). The assay was repeated with the same stocks of bacteria and Caco-2 cells, but the results were similar.



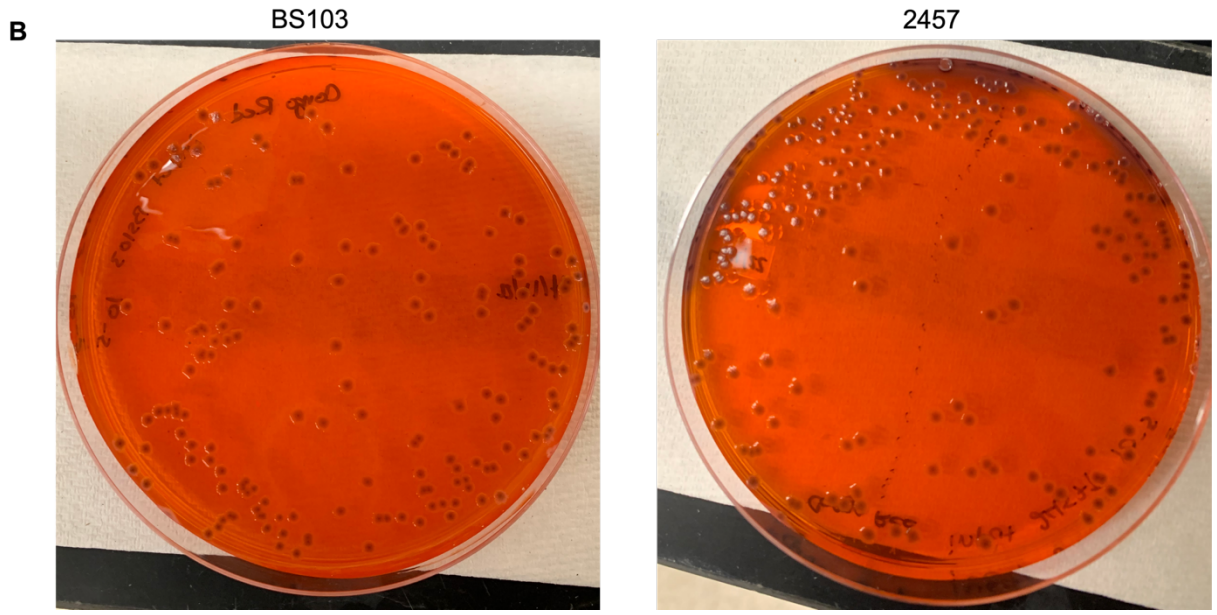


Figure 5-1: ***Plates from the First Attempt to Reproduce the Plaque Assay.*** *A): Resulting plates of BS103 2457T show no plaques after 3 days of incubation. B): Input bacteria for the assay on Congo red plates. Colonies appear to be red with clear or white "halos".*

After these initial results, we ordered fresh Caco-2 cells and virulent *S. flexneri* 2457T from ATCC. We also switched the plates we were using for the assay from six-well plates to 150 mm cell culture dishes, similar to what we use for our input bacteria colonies. Unfortunately, we still were not seeing any plaques forming on the monolayers using these fresh cell lines and bacteria **(Figure 5-2**). In fact, the monolayers appeared to be deteriorating during the assay. This can be seen in **Figure 5-2** as gaps of varying sizes in the monolayers. These can be verified as deteriorating monolayers and not plaques due to the fact that they are present in both 2457T experiment and the BS103 control. We repeated the process but saw similar results each time.

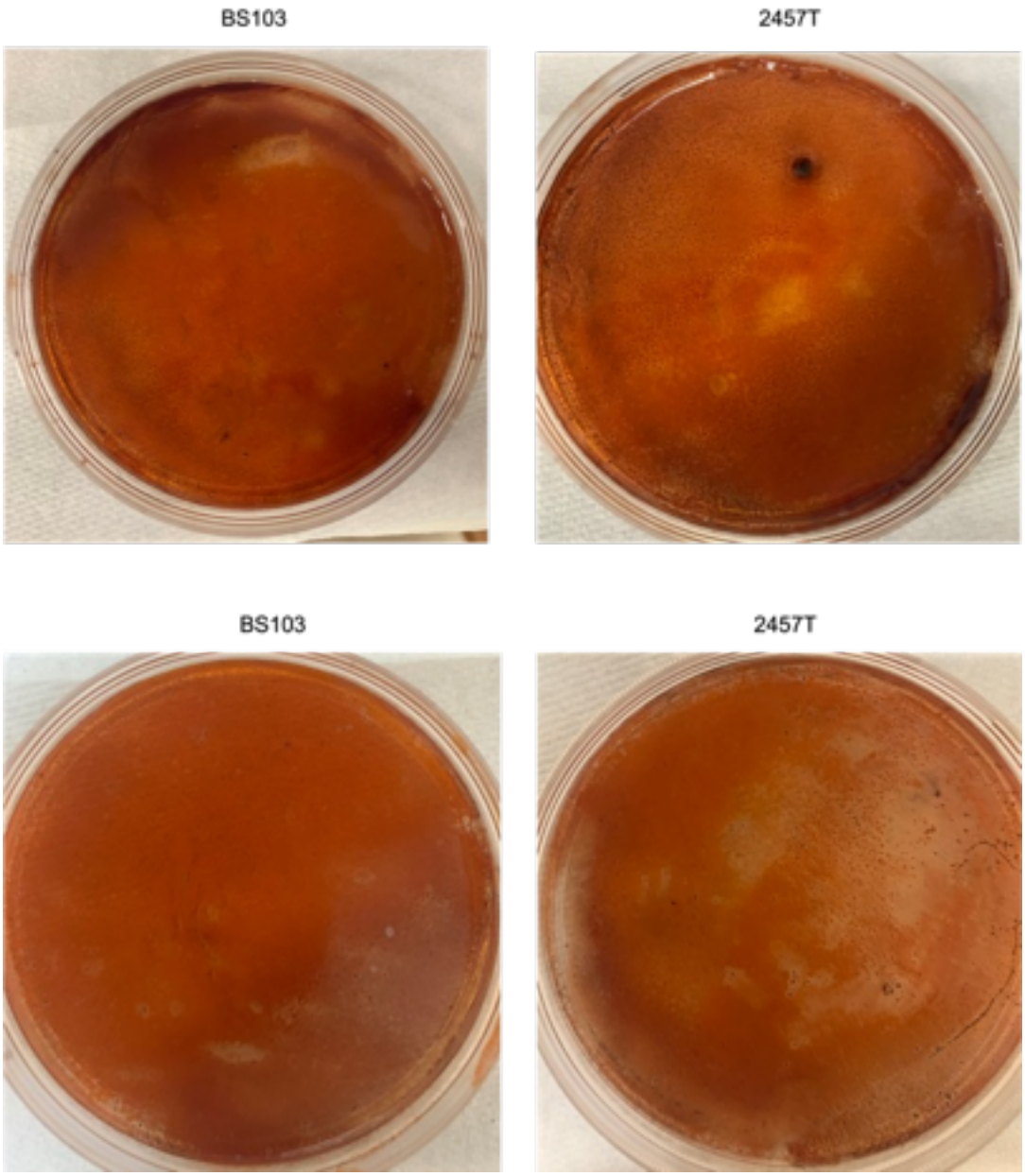


Figure 5-2: ***Tissue Culture Plates After the Plaque Formation Assay.*** *Examples of the Caco-2 monolayers deteriorating before the completion of the assay. Gaps in the monolayers can be seen in both the 2457T experiments and the BS103 controls.*

We wanted to try and grow healthier monolayers before conducting the plaque formation assays. To do this, we tried to grow our Caco-2 cells in RPMI media as well as trying different concentrations of FBS (10%-20%).[3] When we tried to perform a plaque assay with these confluent monolayers, the results were even worse than before. **Figure 5-3** shows a plate where

almost all the cells in the monolayer are gone. After this, we tried the DMEM media again. We thought that it was possible that our monolayers were not reaching the proper confluence, so we let them grow longer before infecting them with *Shigella*. Strangely, we experienced issues getting our tissue cultures to reach high cell density. With the plaque assay not showing success, we checked if our monolayers would survive the gentamicin protection invasion assay. As expected, the monolayers failed to keep their structural integrity throughout the experiment and fell apart during the wash steps.

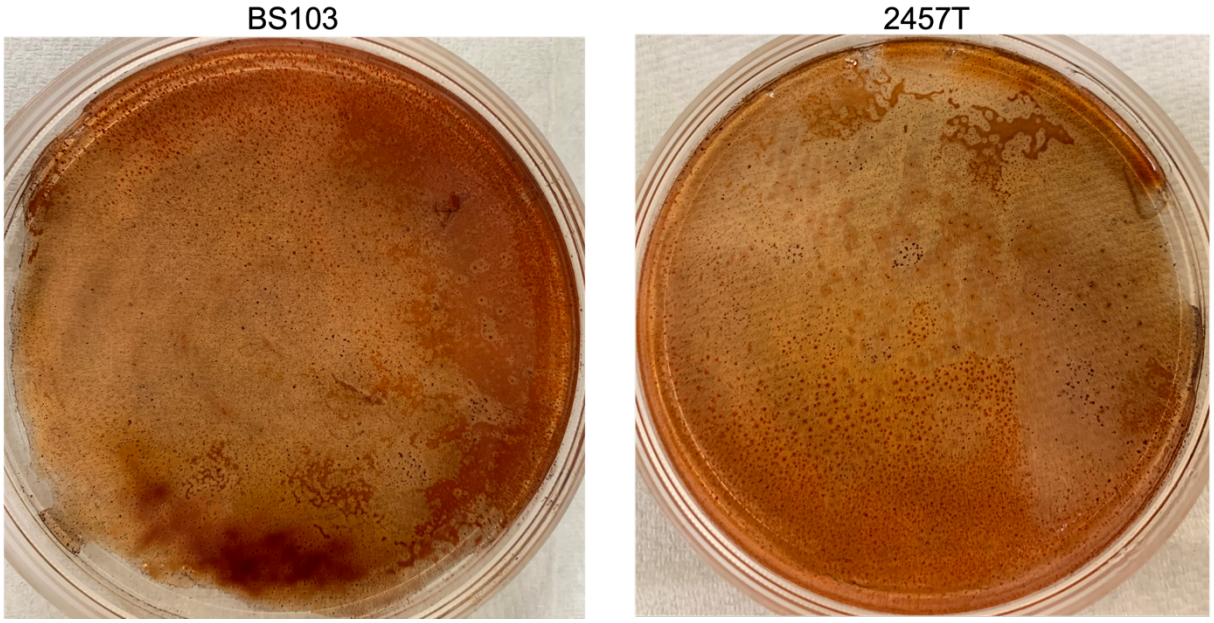


Figure 5-3: ***Deteriorating Monolayers of Caco-2 After Plaque Assay.*** *Monolayers are not surviving the experiment*

Discussion

In the past, our lab has been able to use the plaque formation assay and the gentamicin protection assay to evaluate lead compounds and their ability to inhibit *Shigella* pathogenesis.[2] Here, we wanted to reproduce these experiments to be able to test our hit compounds from the

~1.7 million compound library screen carried out in our collaboration with GSK. The screen was able to identify 44 hit compounds after primary and secondary screens, but reviewers wanted verification of the compounds' activity against intracellular *Shigella*.

The major issue we had in producing an assay to test these compounds was the inability to get our Caco-2 monolayers to survive the rigorous experiment while infected with *Shigella*. We also experienced issues in achieving high cell density, despite different medias and varying growth times. **Figure 5-4** shows the difference between low density and high density Caco-2 monolayers. Most of our experiments were done on monolayers somewhere between the high and low density seen in **Figure 5-4**. In the future, the experiment might have been more successful if we could get the cultures to grow better.

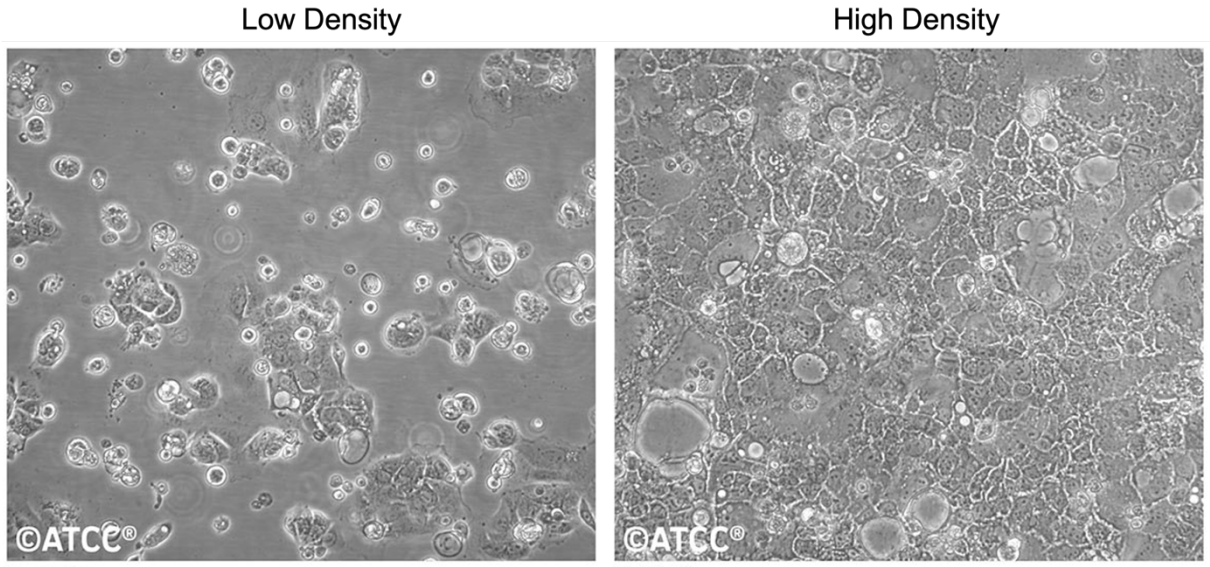


Figure 5-4: ***Examples of Caco-2 Cell Densities by ATCC.***

References

(1) Maurelli, A. T.; Blackmon, B.; Curtiss Iii, R. *Loss of Pigmentation in Shigella Flexneri 2a Is Correlated with Loss of Virulence and Virulence-Associated Plasmid*; 1984; Vol. 43.

(2) Emanuele, A. a; Adams, N. E.; Chen, Y.-C.; Maurelli, A. T.; Garcia, G. a. Potential Novel Antibiotics from HTS Targeting the Virulence-Regulating Transcription Factor, VirF, from Shigella Flexneri. *J. Antibiot. (Tokyo).* **2014**, *67* (5), 379–386. https://doi.org/10.1038/ja.2014.10.

(3) Ferraretto, A.; Bottani, M.; De Luca, P.; Cornaghi, L.; Arnaboldi, F.; Maggioni, M.; Fiorilli, A.; Donetti, E. Morphofunctional Properties of a Differentiated Caco2/HT-29 Co-Culture as an in Vitro Model of Human Intestinal Epithelium. *Biosci. Rep.* **2018**, *38* (2), 1–16. https://doi.org/10.1042/BSR20171497.

Chapter 6 Concluding Remarks

Bacterial infections remain a major disease burden, particularly in regions of the world where access to adequate medication and sanitation are lacking.[1,2] Due to many factors including lack of new antibiotic production and their improper use, antibiotic resistance continues to emerge with many bacteria now becoming multi-drug resistant.[1] Of the current treatments for *Shigella*, resistance to ciprofloxacin or azithromycin has risen to 17% in clinically-isolated strains in the United States.[3] Worldwide, antibiotic resistance in *Shigella* is significantly rising, with "complete ciprofloxacin resistance (MIC ≥ 4 mg/L)" noted as a "Serious Concern" by the *World Health Organization*.[4] Increased incidence of antibiotic resistance in *Shigella*, among other pathogens, and increased mortality highlight a critical need to develop novel antibiotics, vaccines, or other therapies to treat infection. One promising approach to developing improved therapeutics to treat shigellosis, and other infections, is targeting virulence pathways by which the causative pathogens invade and propagate within infected hosts. By targeting virulence rather than cell viability, it is thought that there will be a weaker selective pressure to evolve resistance to the drug.[5] Additionally, targeting virulence pathways should have no effect on the avirulent microbiome thereby reducing the risk of opportunistic infections, e.g., *C. difficile*, which are held in check by some of the avirulent microbes.[5]

We believe VirF is a very promising antivirulence target in *Shigella* for several reasons. VirF is the master transcriptional regulator of the *Shigella* virulence pathway, binding the *virB* promoter leading to the downstream expression of *ipaB*, *ipaC*, and *ipaD*, as well as other genes that encode for structural and effector proteins that are key for infection.[6–11] VirF also activates

transcription of *icsA*, initiating cell-to-cell spread by facilitating the polymerization of host cell actin at one pole of the bacterium.[10,11] Additionally, VirF is only expressed under conditions found in the host colonic lumen, which makes it unlikely that resistance will develop due to environmental antibiotic exposure.[5] The purpose of the studies in this dissertation was to gain insight on how the VirF DNA binding domain interacts with the *virB* promoter and it's potential as a antivirulence target.

AraC proteins such as VirF have been difficult proteins to study.[12] We set out to develop a more efficient method to express and purify native VirF and the C-terminal DNA binding domain of VirF. Although we fell short of obtaining active VirF in a useful yield, we believe our results could still act as a reference for future attempts to study this protein. Due to the lack of structural information on VirF, we used two *E. coli* homologs, GadX and MarA, to generate homology models of the VirF DNA-binding domain (DBD). This allowed us to study the interactions between VirF and the *virB* promoter in comparison with MarA binding to the *marRAB* promoter. This gave insight to how the VirF DBD binds its DNA and how this interaction may be exploited as an antivirulence drug target. We also tried to develop chimeric proteins, swapping portions of the VirF binding domain with MarA. Our objective was to use a chimeric form of MarA to further study this interaction, and to further investigate our potential inhibitor, compound 19615.[13,14] Even though we could not induce binding to the VirF promoters by the chimeric protein, we believe this interaction is still an attractive target for antivirulence inhibition. With more time, this interaction could be further probed by designing a small library of MarA-VirF chimeric proteins. Considering the specificity and complexity of the protein-DNA interaction, several minor yet specific changes may be needed to identify an acceptable mutant to use in these studies. In conclusion, we believe these experiments will lead to an

improved understanding of not only the VirF•*virB* and the MarA•*marRAB* interaction, but hopefully add to the comprehension of the AraC family of transcriptional activators.

References

(1) Stevens, P. Diseases of Poverty and the 10/90 Gap. In *Fighting the Diseases of Poverty*; 2004.

(2) Médecins Sans Frontières. *Fatal Imbalance: The Crisis in Research and Development for Drugs for Neglected Diseases*; 2001.

(3) Centers for Disease Control and Prevention. Antibiotic Resistance Threats in the United States. *U.S. Dep. Heal. Hum. Servies* **2019**, 1–113. https://doi.org/10.15620/cdc:82532.

(4) Williams, P.; Berkley, J. A. Dysentery (Shigellosis) Current WHO Guidelines and the WHO Essential Medicine List for Children http://www.who.int/selection_medicines/committees/expert/21/applications/s6_paed_antibiotics_appendix5_dysentery.pdf.

(5) Clatworthy, A. E.; Pierson, E.; Hung, D. T. Targeting Virulence: A New Paradigm for Antimicrobial Therapy. *Nat. Chem. Biol.* **2007**, *3* (9), 541–548. https://doi.org/10.1038/nchembio.2007.24.

(6) Gallegos, M. T.; Schleif, R.; Bairoch, A.; Hofmann, K.; Ramos, J. L. Arac/XylS Family of Transcriptional Regulators. *Microbiol. Mol. Biol. Rev.* **1997**, *61* (4), 393–410. https://doi.org/9409145.

(7) Martin, R. G.; Rosner, J. L. The AraC Transcriptional Activators. *Curr. Opin. Microbiol.* **2001**, No. 4, 132–137.

(8) Tobe, T.; Yoshikawa, M.; Mizuno, T.; Sasakawa, C. Transcriptional Control of the Invasion Regulatory Gene VirB of Shigella Flexneri: Activation by VirF and Repression by H-NS. *J. Bacteriol.* **1993**, *175* (19), 6142–6149.

(9) Adler, B.; Sasakawa, C.; Tobe, T.; Makino, S.; Komatsu, K.; Yoshikawa, M. A Dual

Transcriptional Activation System for the 230 Kb Plasmid Genes Coding for Virulence-Associated Antigens of Shigella Flexneri. *Mol. Microbiol.* **1989**, *3* (5), 627–635. https://doi.org/10.1111/j.1365-2958.1989.tb00210.x.

(10) Sakai, T.; Sasakawa, C.; Yoshikawa, M. Expression of Four Virulence Antigens of Shigella Flexneri Is Positively Regulated at the Transcriptional Level by the 30 Kilo Dalton VirF Protein. *Mol. Microbiol.* **1988**, *2* (5), 589–597. https://doi.org/10.1111/j.1365-2958.1988.tb00067.x.

(11) Bernardini, M. L.; Mounier, J.; D'Hauteville, H.; Coquis-Rondon, M.; Sansonetti, P. J. Identification of IcsA, a Plasmid Locus of Shigella Flexneri That Governs Bacterial Intra- and Intercellular Spread through Interaction with F-Actin. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86* (10), 3867–3871. https://doi.org/10.1073/pnas.86.10.3867.

(12) Schleif, R. AraC Protein: A Love-Hate Relationship. *BioEssays* **2003**, *25* (3), 274–282. https://doi.org/10.1002/bies.10237.

(13) Emanuele, A. a; Adams, N. E.; Chen, Y.-C.; Maurelli, A. T.; Garcia, G. a. Potential Novel Antibiotics from HTS Targeting the Virulence-Regulating Transcription Factor, VirF, from Shigella Flexneri. *J. Antibiot. (Tokyo).* **2014**, *67* (5), 379–386. https://doi.org/10.1038/ja.2014.10.

(14) Emanuele, A. A.; Garcia, G. A. Mechanism of Action and Initial, In Vitro SAR of an Inhibitor of the Shigella Flexneri Virulence Regulator VirF. *PLoS One* **2015**, *10* (9), e0137410. https://doi.org/10.1371/journal.pone.0137410.