7-2020

Structural Analysis of the Multifunctional SpoIIE Regulatory Protein of Clostridioides difficile.

Blythe Emily Bunkers
University of Arkansas, Fayetteville

Structural Analysis of the Multifunctional SpoIIE Regulatory Protein of *Clostridioides difficile*.

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Cell and Molecular Biology

by

Blythe Emily Bunkers
University of Arkansas
Bachelor of Science in Biology, 2018

July 2020
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

_____
Mack Ivey, Ph.D.
Thesis Director

_____
Timothy Kral, Ph.D.
Committee Member

_____
Daniel Lessner, Ph.D.
Committee Member

_____
Alejandro Rojas, Ph.D.
Committee Member

ABSTRACT

*Clostridioides (*formally *Clostridium*) *difficile* is a medically relevant pathogen pertinent to infectious disease research. *C. difficile* is distinctly known for its ability to produce two toxins, enterotoxin A and cytotoxin B, and the propensity to colonize the mammalian gastrointestinal tract. It is known that metabolism is tightly correlated with sporulation in endospore producers such as *C. difficile*, but an interesting and novel regulatory relationship found by the Ivey lab has yet to be understood. The relationship explored in this study is observed between the sporulation factor, SpoIIE, which represses expression of an ABC peptide transporter, app. In this study, two primary approaches were taken in order to investigate this interaction. The first method involved directly mutating the gene for SpoIIE in attempt to understand which portion of the gene is crucial for retaining protein functionality, in this case, the repression of app. The two genes of interest were readily available on their own individual plasmids, which easily allowed for mutagenesis experimentation before transformation into *E. coli*. The plasmid with app was constructed to include a fluorescent probe, allowing the relative level of repression to parallel the qualitative measurement of fluorescence. On the SpoIIE containing plasmid, missense point mutations, done using PCR site-directed mutagenesis, and a large deletion of the SpoIIE protein transmembrane anchor, done by endonuclease restriction digest, were performed in pursuit of this investigative method. The secondary approach was to computationally model the SpoIIE protein structures resultant of the genetic mutations done in vitro. Side-by-side images of the predicted mutant models produced from the experimentally preferred algorithm were compared to the wildtype SpoIIE model. The visual analysis of the structures and the comparison of various algorithms is anticipated to be insightful for not only this project, but future exploration into SpoIIE as a multifunctional protein within *C. difficile*.

ACKNOWLEDGMENT

I would like to thank the University of Arkansas, the Department of Biological Sciences, and the Cell and Molecular Biology Program for providing my graduate funding through the form of a teaching assistantship.

My deepest thank you goes to my master's thesis advisor, Dr. Mack Ivey, for accepting me into his lab, as well as his eagerness to share his seemingly endless knowledge on this topic of research. His dedication to all of his students' success is admirable and deeply appreciated.

My special thanks to Dr. Elizabeth Padilla-Crespo for her mentorship, training, and expertise. From the very beginning, she instilled a solid foundation of molecular lab skills upon which my research is built.

I would also like to thank my thesis committee members: Dr. Kral, Dr. Lessner and Dr. Rojas for their range expertise and assistance. They have unconditionally offered support and advice throughout my time as a student at the University of Arkansas.

I am so grateful to all my lab members, both past and present. Thank you to my current lab member Danielle Maynard for her initial training on this project and friendship. Thank you to my previous lab members and colleagues who have in turn become wonderful, irreplaceable friends: Kelsey Bevenour, Atem Forwang and many others. These friendships provided me with a vital support system throughout the progression of my thesis work.

Finally, I must express my gratitude to my family for their love and support which made this great feat possible. Thank you to Donna Bunkers for hour-long phone calls packed with advice, Mel Bunkers for his rational and calming demeanor, and Samuel Bunkers for making me laugh every bit of the way. Lastly, thank you to Daniel Lawson for his dedication, patience and love through years of stress. There is no way to aptly express my gratitude through written word.

I would also like to acknowledge many other individuals unlisted, as my network of support in both friends and family members is far too many to include each name. I am beyond grateful to have too plentiful of a list.

TABLE OF CONTENTS

ABBREVIATIONS

AAD	antibiotic associated diarrhea
AA PMC	antibiotic associated pseudomembranous colitis
AS	active surveillance
Cb	carbenicillin
CDI	*C. difficile* infection
CFU	colony forming units
DK	aspartic acid → lysine mutant
DQ	aspartic acid → glutamine mutant
GFP	green fluorescent protein
GMQE	global model quality estimate
HA	hospital associated
Kb	kilo-basepairs
Km	kanamycin
LB	Luria Bertani growth media
MDROs	multidrug-resistant organisms
MRSA	methicillin-resistant *Staphylococcus Aureus*
NMR	nuclear magnetic resonance
Opp	oligopeptide permease protein
PCR	polymerase chain reaction
PDB	protein database
QSQE	quaternary structure quality estimate
RT	room temperature
TM	transmembrane
UA321	University of Arkansas *E. coli* strain #321 (applies to all)
pUA321	plasmid isolated from University of Arkansas strain #321 (applies to all)

I. INTRODUCTION

I.1. Infectious Diseases: an overview

Infectious diseases are often thought of as agents with unknown origins infiltrating the civilized world, but in actuality, "Infectious Disease" is a term that describes a vast number of pathogens. The study of infectious diseases involves a complex network of dynamics observed between the host and parasite and is therefore inherently multidisciplinary (Engering et al., 2013). This field of research relies on biologists, ranging from molecular and microbial to ecological and clinical, in order to develop a comprehensive understanding of an infection. In any clinical context, whether that be pharmaceutical treatment, epidemiological trends, or prevention medicine, the fundamental three-way interactions between the host, parasite and their physical environment must be considered.

Infectious disease research is centralized around the relationship between the parasite and its host. Parasitism is an essential biological relationship and one of the most ubiquitous life strategies on Earth, as all living organisms are parasitized by at least one other species (Albert et al., 2002; Balloux & van Dorp, 2017). Parasite is a broad term that can be further divided into five major types: helminth, fungus, protozoan, bacteria, and viruses (Albert et al., 2002; Janeway et al., 2001; Kilpatrick & Altizer 2010). Prion is occasionally listed as a sixth category depending upon the source, as they are infectious proteins but not parasites (Albert et al., 2002; Barreto et al., 2006). More specifically, a *pathogen* is a parasite that causes disease to its host, whereas by comparison few parasites co-exist symbiotically with their hosts in the absence of disease (Balloux & van Dorp, 2017; Barreto et al., 2006; Casadevall & Pirofski 2000; Kilpatrick & Altizer 2010). In order to survive, pathogens must colonize a host, subvert both innate and adaptive immune responses, replicate, and lastly… spread to subsequent hosts (Albert et al., 2002; Janeway et al., 2001).

By definition, disease describes a disorder in structure or function, especially alongside the production of explicit and identifiable symptoms (Balloux & van Dorp, 2017; Kilpatrick & Altizer 2010). Often times, disease symptoms presented by an infection aid in the transmission to the next host (Albert et al., 2002). For example, the common cold, caused by the rhinovirus, produces excess mucous along with sneezing and coughing; actions that aim to aerosolize viral particles in hopes of inhalation by another host to exploit (Jacobs et al., 2013). Many diseases present in the human population are not classified as infectious, but are rather inherited or congenital, one of the most common being cardiovascular disease (Mc Namara et al., 2019). Contrastingly, infectious diseases are caused by transmissible and communicable pathogens (Barreto et al., 2006). This defining quality of being "transmissible" allows for the analysis of how pathogens are maintained and transmitted within a host population, which often times have significant clinical implications. Again, emphasizing that understanding these large-scale infection dynamics is vital and parallel to the analysis of molecular interactions.

Not all infectious diseases arise from the untamed wilderness in spillover events. In fact, the majority of human pathogens live and thrive in the same environment as us; their hosts (Dunn et al., 2010). Common human pathogens can be isolated from nearly any public space. The outdoor environment is home to numerous potential pathogens in the soil such as *Toxoplasmosis gondii* and *Pseudomonas aeruginosa*, while streams can be home to microbes such as *Giardia lamblia* and *Legionella pneumonia* (Gao et al., 2016; Green et al., 1974; McClung et al., 2017). Raw and unwashed food found in the produce section from a market is an ideal environment for *Escherichia coli* or *Listeria monocytogenes* (Zeng et al., 2014; Zhang et al., 2018). Industrial facilities, such as a sewage treatment plants, are often the environment for pathogens such as *Enterococcus faecalis* and *Proteus vulgaris* (Cyprowski et al., 2018). Even

medical settings, such as hospital rooms (places optimistically thought to be sterile) prove to be a niche environment for opportunistic pathogens such as methicillin-resistant *Staphylococcus Aureus* (MRSA) and *C. difficile* (Hassoun et al., 2017; Walters et al., 1983). Pathogens are evidently omnipresent, but their pervasiveness is chiefly controlled by host availability and transmission routes (Kilpatrick & Altizer 2010).

The prevalence of all pathogens is largely dependent on and limited by their mode of transmission (Antonovics et al., 2017; Barreto et al., 2006; Dunn et al., 2010). There are five modes of pathogen transmission: direct contact, inhalation, fecal-oral, vector-borne and fomite (Antonovics et al., 2017). Direct contact infections occur through an open wound or mucous membrane, such as with sexually transmitted diseases like *Chlamydia trachomatis* (Phillips 2019). Inhalation transmission is seen in respiratory viruses like coronaviruses, or SARS-CoV-2 (Kucharski et al., 2020). The fecal-oral transmission route is utilized by pathogens like *Vibrio cholerae* or *Salmonella enterica;* the culprits behind many cases of food poisoning (Diard & Hardt 2017; Rinaldo et al., 2017). Vector-borne pathogens are transmitted to hosts via arthropods like ticks and mosquitos, some of these being *Borrelia burgdorferi*, *Plasmodium falciparum* and all *flaviviruses* (Rossati et al., 2016; Schorderet-Weber et al., 2017; Leitner et al., 2015). The fifth mode, fomite, uniquely requires a secondary mode of transmission, such as ingestion (Antonovics et al., 2017). The term fomite refers to objects likely to carry microbes that have the ability to persist in the environment for prolonged periods of time in the absence of a host (Alberts et al., 2002). Endospore producers commonly exist in their dormant spore-form in the environment on a fomite, and then germinate into vegetative cells once being either ingested or inhaled by a susceptible host (Deakin et al., 2012; Saujet et al., 2013). *C. difficile* is an archetypal

pathogen that employs this evolutionarily advantageous method of transmission (Paredes-Sabja et al., 2014).

In the case of microparasites, such as bacteria, the host can additionally be thought of as the "environment" (Kilpatrick & Altizer 2010). With the exception of fomite transmission, most parasites cannot survive extensively in the exterior environment without a host. Parasites are entirely dependent upon their hosts for protection and nutrients in order to subsist (Albert et al., 2002; Kilpatrick & Altizer 2010). Gastrointestinal pathogens, transmitted through the fecal-oral route, are ideal examples when considering how the host can double as the environment.

The human body's temperature, in combination with an anaerobic setting and a radical drop in pH due to hydrochloric acid found in the stomach, create a unique environment for invading intestinal pathogens to endure (Fimlaid et al., 2015). Furthermore, as many studies have published, the human microbiome is a notably complex ecosystem within itself, as the colon is home to upwards of five-hundred bacterial species alone (Jandhyala et al., 2015; Lozupone et al., 2012). This intestinal environment proves to be a tricky one for attacking pathogens as they face high temperatures, little oxygen, an acidic pH, and the normal gut flora as powerful resource competitors (Lozupone et al., 2012). The successful invaders of the gastrointestinal tract have shown to be both opportunistic and resilient.

Opportunistic pathogens are defined in medical literature as organisms that "become pathogenic following a perturbation to their host" (Brown et al., 2012). Nosocomial infections, ones originating from hospitals, are conventionally and fundamentally understood to be driven by opportunistic pathogens (Fimlaid et al., 2015). Hospitals are home to the weakest and most susceptible host population while simultaneously housing some of the most aggressive and resilient pathogens. Numerous hospital residents are recipients of heavy antibiotics following a

range of routine procedures (Rupnik et al., 2009). In many cases, these antibiotic therapies protect patients against subsequent infections, but inadvertently, this makes patients intestinally available for opportunistic colonization as normal gut flora become deficient (Barreto et al., 2006; Brown et al., 2012). Due to their unique infection dynamics within a specific niche, an entire division of infectious disease research is dedicated to nosocomial infections (Alberts et al., 2002; Casadevall & Pirofski 2000). *C. difficile* is the ideal model of an opportunistic, nosocomial pathogen as these intestinal infections almost exclusively originate from healthcare settings (Deneve et al., 2009; Dodson & Borriello 1996).

Aside from being opportunistic, *C. difficile* is also considered to be a highly resilient pathogen because of its ability to produce endospores (Deakin et al., 2012; Fimlaid et al., 2015). As briefly mentioned, the production of endospores allows for *C. difficile* to withstand a variety of harsh environmental factors. Spores can exist on various medical surface fomites as they are resistant to desiccation and disinfectants (Lawley et al., 2009). These metabolically dormant spores can then be unintentionally ingested by abundantly available and susceptible hosts (Paredes-Sabja et al., 2014). Not only do the spores successfully evade death by disinfectants, they also have the uncanny ability to survive the stomach pH, and then germinate in the intestinal tract (Fimlaid et al., 2015). Successful colonization of the intestinal tract by *C. difficile* is primarily seen in the absence of resource competitors, or the normal gut microbes removed through antibiotic therapy (Walters et al., 1983).

The aforementioned infection dynamics are multifaceted and highly distinctive to this system. The transmission route, niche susceptible host population, and evolutionary reproductive tactics, make *C. difficile* a fascinating organism in which to study infection at a molecular scale.

C. difficile is a medically relevant pathogen with unique microbial properties that make it an interesting model for infectious disease studies.

I.2. *C. difficile*: history and facts

C. difficile is a gram positive, anaerobic, endospore producing bacilli that is virtually endemic to hospitals and their population of vulnerable hosts (Redelings et al., 2007). In the context of this infectious disease, a vulnerable host is one who has deficient normal intestinal flora or a weakened immune system (Bartlett & Gerding 2008). Generally, these are characteristics resultant of both broad-spectrum antibiotics and chemotherapy, two commonly received medical treatments, which have aided in the creation of a niche population of susceptible hosts for *C. difficile* in the medical setting (Murphy et al., 2012; Redelings et al., 2007). Furthermore, transmission within the hospital patient host population, demonstrates horizontal and density-dependent transfer between individuals (McDonald 2013). Horizontal transmission indicates that any individual within the identified host population can become infected, while density-dependent refers to the statistical increase in incidence seen with an increase in available hosts (Kilpatrick & Altizer 2010; McDonald 2013).

C. difficile is the reigning cause of nosocomial gastrointestinal diseases and a leading causative agent for antibiotic associated diarrhea (AAD) (15-25% of all reported cases) in Western and European countries for nearly five decades (Bouza et al., 2012; Dubberke & Olsen 2012; Freeman et al., 2010). In the United States alone, the *C. difficile* mortality rate reported an increase of 5.7 per million in 1999, to a rate of 23.7 per million just five years later in 2004 (Redelings et al., 2007). *C. difficile* remains a prevalent nosocomial pathogen because of its ability to evade antimicrobials through endospore production, seen both inside the host and in the exterior environment (Fimlaid et al., 2015). Despite being anaerobic, *C. difficile* persists without

a host in an oxygenic environment by existing as spore fomites (Fimlaid et al., 2015; Pereira et al., 2013; Underwood et al., 2009). These infectious spores survive surface disinfectants and germinate into vegetative cells after ingestion (Casadevall & Pirofski 2000; Janeway et al., 2001). This is representative of the fecal-oral transmission route as discussed above. Within the host, that same ability to produce spores also contributes to the survival of *C. difficile* despite the use of antibiotics aimed at clearing the infection (Walters et al., 1983).

The disease spectrum seen in *C. difficile* infections is highly variable (Bartlett & Gerding 2008; Deneve et al., 2009). Symptoms range from mild diarrhea and colitis, to severe toxic megacolon and sepsis (Bartlett & Gerding 2008; Carman et al., 2011). The most severe cases result in antibiotic-associated pseudomembranous colitis (AA PMC), a symptom observed only after established colonization in the colon of patients under heavy antibiotic therapy (Lawley et al., 2009; Walters et al., 1983). PMC often presents with accompanying symptoms such as intestinal distention, colitis, abdominal pain, and severe diarrhea (Reinke & Messick 1994; Surawicz & McFarland 1999). If improperly treated, PMC can further progress into toxic megacolon and present complications such as organ failure and sepsis (Borriello, 1990; Dodson & Borriello, 1996; Surawicz & McFarland, 1999; Deneve et al., 2009). As seen with most pathogens, the course of infection produces symptoms or altered "host behavior" that aim to increase the likelihood of transmission (Antonovics et al., 2017; Casadevall & Pirofski 2000). The pathogenicity of *C. difficile* results in heavy diarrhea and spore shedding into the environment; a host behavior change that propagates the pathogen's spread (Engering et al., 2013). Not only does this elucidate how *C. difficile* dominates as a resilient and opportunistic nosocomial pathogen, but this emphasizes the importance of preventative measures for health

care workers (i.e. gloves, handwashing, etc.) in order to thwart successive spreading to additional patients (Kampf et al., 2009; Lessa et al., 2012; Paredes-Sabja et al., 2014).

Antibiotic resistance is a growing topic of concern, particularly for nosocomial pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA) and *C. difficile* (Lawley et al., 2009; Walters et al., 1983). Their overexposure to antimicrobials (environmental disinfectants and broad-spectrum antibiotics) has led to their adaptation to withstand such treatments and the ensuing emergence of highly virulent strains (Deneve et al., 2009; Fawley et al., 2007; Lawley et al., 2009). Multi-drug resistant organisms (MDROs) no longer respond to traditionally effective medications which poses a substantial limitation for patient care and treatment (Walters et al., 1983). Since *C. difficile* is already an aggressive and invasive pathogen, there are ongoing programs that monitor the existence and prevalence of specific strains, as well as precautions that are taken by health professionals to prevent or delay the appearance of hyper-virulent strains (Lessa et al., 2012; Walters et al., 1983). One being the intentional avoidance of vancomycin (a powerful broad-spectrum antibiotic) to treat *C. difficile* patients in order to preclude the development of vancomycin-resistant strains (Surawicz & McFarland, 1999; Walters et al., 1983). Not only does *C. difficile* exhibit advantageous transmission tactics that make it incredibly successful in healthcare settings, but its pathogenicity and ability to produce toxins play a major role in overall virulence (Di Bella et al., 2016; Lessa et al., 2012; Paredes-Sabja et al., 2014).

I.3. Pathology: metabolism and toxicity

After accidental ingestion by a susceptible host, *C. difficile* spores resist destruction by gastric acid and germinate after reaching the small intestines (Lessa et al., 2012; Paredes-Sabja et al., 2014). Once in its vegetative and metabolically active state, *C. difficile* colonizes mammalian intestines in the absence of normal gut flora, typically after long term antibiotic usage (Depestel

& Aronoff 2013; Lessa et al., 2012). With the lack of resource competitors, *C. difficile* proliferates and dominates the intestinal microbial population (Depestel & Aronoff 2013). Following successful colonization, *C. difficile* adheres to intestinal epithelial cells and produces two toxins which heavily influence the course of the infection (Borriello 1990; Borriello et al., 1990). These toxins play an insidious role during a *C. difficile* infection and a plethora of studies have addressed these toxins, their structures, and the devious effect they have on the host (Depestel & Aronoff 2013; Di Bella et al., 2016; Lessa et al., 2012; Taylor et al., 1981).

The ability to adhere to the intestinal epithelium acts as an accessory virulence factor for *C. difficile* as it aids in toxin delivery and penetration (Di Bella et al., 2016; Karjalainen et al., 1994). The destructive virulence factors underlying a *C. difficile* infection are two large exotoxins, toxin A and B, encoded by genes *tcdA* and *tcdB* within the chromosomal DNA pathogenicity locus (*paloc*) (Di Bella et al., 2016; Lessa et al., 2012). Toxin B was the first to be discovered and was initially named "cytotoxin" (Taylor et al., 1981). TcdB is a 250-270 kd cytotoxin that leads to depolymerization of actin and other cytoskeletal components in mammalian cell lines in vitro, resulting in the demise of cellular structure and function (Di Bella et al., 2016; Hecht et al., 1988; Ottlinger & Lin 1988). Toxin A is a 308 kb enterotoxin that tightly binds epithelial cell receptors, leading to fluid secretion, inflammation, and tissue necrosis (Just et al., 1995; Kelly et al., 1994**)**. Additionally, tcdA & tcdB glycosylate small Rho proteins using UDP-glucose as the sugar donor and lead to cytoskeleton disorganization (Deneve 2009) Among common strains, toxin production increases as *C. difficile* enters the stationary growth phase, whereas hypervirulent strains also interestingly display toxin production during the exponential phase (Deneve et al., 2009; Freeman et al., 2010).

Being that *C. difficile* is an endospore forming bacteria, sporulation is another metabolic event also influencing the pathogen's virulence. Sporulation occurs when vegetative growth is no longer metabolically favored, and both environmental and cellular signals initiate a complex cascade of activations ultimately leading to the production of spores (Edwards & McBride 2014; Saujet et al., 2013; Underwood et al., 2009). Epidemic strains of *C. difficile*, such as ribotypes 001 and 027, even demonstrate an increase in sporulation capacity, reflected in their increased virulence (Fawley et al., 2007; Freeman et al., 2010). Spore production is controlled through the activation of sporulation factors which govern transcription factors, and ultimately up-regulation of the genes responsible for spore formation (Edwards & McBride 2014; Fimlaid et al., 2013; Paredes-Sabja et al., 2014; Pereira et al., 2013).

The pathogenicity observed in *C. difficile* is multifaceted as it involves dynamics that must take into account metabolic processes such as spore formation and toxin production, both of which are direct links to the organism's virulence. Metabolism in general is tightly correlated with the availability of nutrients and resource utilization (Casadevall & Pirofski 2000). As *C. difficile* is living in a peptide rich environment, the gastrointestinal tract, it can be inferred that successful interaction with the intestinal environment and subsequent cellular transport of peptides is essential for colonization.

I.4. The relationship between SpoIIE and ABC peptide transporter, app

In previous experiments done by the Ivey lab, screens of genomic libraries were preformed in order to identify gene products regulating peptide transport in *C. difficile*. Studying cellular peptide transport can be insightful for understanding resource utilization and metabolism in the context of virulent organisms, which is why the preceding project was pursued. The central method used in these screenings involved the use of a green fluorescence protein (GFP),

which was fused onto the promoter region of the app (formerly opp) oligopeptide transporter. This allowed for the relative repression levels to be viewed as relative brightness, as the GFP gene, under the control of the *app* promoter, was repressed. Two proteins, MtlF and SpoIIE, were identified as transcriptional repressors based upon their reduced amount of fluorescence after screening, and investigating SpoIIE became the focal point of this research project.

SpoIIE is a sporulation factor found in all gram-positive, spore-forming bacteria (Saujet et al., 2013). It is most aptly characterized in the model organism, *B. subtilis*, which is analogous to the *C. difficile* model discussed in this study. SpoIIE is a multifunctional protein directly involved in the regulation and timing of spore development through septum formation and activation of transcription factors (Barak et al., 1996; Carniol et al., 2005). SpoIIE has a TM domain of 10 helices on the N-terminus, a C-terminus phosphatase domain, and a central domain with ambiguous function (Arigoni et al., 1999; Carniol et al., 2005). The TM domain allows for SpoIIE membrane integration and localization, while the cytoplasmic phosphatase domain both catalyzes the dephosphorylation of anti-sigma factor SpoIIAA and also binds specific cytoskeletal proteins such as FtsZ and RodZ (Ben-Yehuda & Losick 2002; Lucet et al., 2000). SpoIIE exists as a dimer, and the dimerization of this protein is believed to play a role in the regulation and timing of the phosphatase domain activity (Arigoni et al., 2002).

The C-terminus portion of SpoIIE, or the phosphatase domain, has previously been crystallized and is more thoroughly studied than the remaining domains of SpoIIE as it belongs to a large family of PP2C serine phosphatases (Arigoni et al., 2002; Barak et al., 1996). In addition to the phosphatase domain, this cytoplasmic region contains a long alpha-helix which serves as the dimer interface while also indirectly interacting with the phosphatase domain via a helical hairpin referred to as the “switch” (Arigoni et al., 2002). The switch rotates upon

dimerization and activates the adjacent phosphatase region of SpoIIE (Arigoni et al., 2002; Carniol et al., 2005). Being that only the C-terminus of the *B. subtilis* SpoIIE protein has been crystallized, there is still much to be determined about the protein structure and function within the central and N-terminus domains, especially when directing this study back to the *C. difficile* model.

The peptide transporter, app, is an oligopeptide permease protein falling into the broad classification of an ABC transporter and ATP-binding cassette, which actively transports substrates such as peptides (Beis 2015; Locher 2016). App consists of five subunits, A, B, C, D and F, each with a specific role in the recognition, binding and ultimately transport of peptides (Steglich et al., 2018). Furthermore, each subunit of app has its own corresponding gene within the app locus (Arigoni et al., 2002). Curiously, the genes are divided and controlled by two promoters in opposite directions (Figure 1.1.). These promoters were mapped previously in the Ivey Lab. Studies with GFP fused to the promoter-containing locus revealed that SpoIIE represses appDF expression. The directness of the connection between SpoIIE expression and repression of appDF is unknown, and the first step to gathering an understanding of this regulatory relationship relies on investigating SpoIIE protein structure and function. This

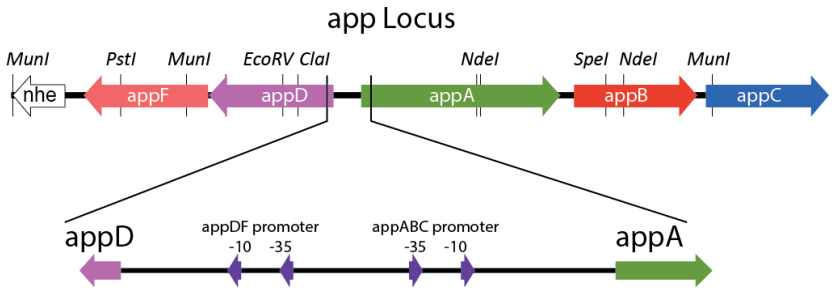

Figure 1.1. App locus Image generated using Adobe Illustrator to edit a VectorNTI

investigation is designed around evaluating mutations within critical domains of SpoIIE in order to assess the affects they have on function.

Within the *B. subtilis* model, specific SpoIIE mutants have been created that are defective in one or all functions mentioned (Barak et al., 1996). Mutations targeting the switch exhibit defective phosphatase activity, while mutations targeting the alpha-helix prevent dimerization and therefore all subsequent protein function. Similar mutations have been targeted within the *C. difficile* SpoIIE protein sequence in order to understand the repression of appDF transcription discussed above. It is unknown whether the repression is direct, mediated, or whether phosphatase activity is required for repression to occur. The following project outlines the experimental design, both in vitro and in silico, and discusses the findings, their implications, and future research directions.

I.5. Research aims

In an attempt to understand the regulatory interaction observed between sporulation factor, SpoIIE, and the ABC peptide transporter, app, specific mutations can be made to SpoIIE in order to genetically investigate which portion of the gene is crucial for maintaining the regulatory function it has on appDF. This is an organized, molecular investigation into a specific genetic relationship that underlies essential pathogen characteristics such as metabolism, sporulation, toxin production, resource utilization and environment interaction. All of these features play a deliberate role in how *C. difficile* interacts with its host and environment and can therefore provide insight into the nature of this infectious disease.

The following is the central hypothesis: If certain mutations within the gene for SpoIIE prove to be non-conservative and demonstrate a loss of protein structure and function, then it can be understood as to which specific portion of SpoIIE is essential for this regulatory interaction with peptide transporter, app.

In order to test this hypothesis, two principal research aims were pursued by implementing molecular genetics methods both in vitro and in silico.

Research Aim 1. The first aim focuses on the genetic investigation of the regulatory properties of SpoIIE and what parts of the gene encoding SpoIIE, and the protein itself, are necessary for retaining this function. This aim was pursued using classical molecular biology techniques to mutate the SpoIIE gene and assess the subsequent protein functionality via fluorescence assay.

Research Aim 2. For the second aim, bioinformatics software was utilized to supplement work done in vitro and build computational models of SpoIIE. This research aim is important because it allows for a visual analysis of SpoIIE protein structure following the mutagenesis experimentation outlined in the first research aim.

II. GENETIC MUTANTS OF SPOIIE AND SCORING OF FUNCTIONALITY

II.1. Introduction

It is known that sporulation is coupled with metabolic pathways such as toxin production, but it remains to be understood why when expressed, this sporulation gene SpoIIE, represses the expression of the gene encoding peptide transporter, appDF, in *C. difficile* as was found by the Ivey lab. The correlation between active resource utilization and the metabolic event of sporulation is crucial to understanding this infectious disease and needs to be examined in the context of overall pathogenicity of *C. difficile*. This curious relationship has yet to be investigated and could elucidate pivotal virulence factors for this destructive and toxic pathogen.

Classical mutagenesis experiments were carried out in order to analyze which portion of the SpoIIE protein is crucial to retain the regulatory function it has on app expression. The two genes of interest were originally isolated from *C. difficile* strain VPI 10463, before transforming into *E. coli* (NEB 5-alpha) for subsequent experimentation. Mutants of the SpoIIE gene were created by either endonuclease restriction digest or by PCR site-directed methods. The two genes of interest, appDF and SpoIIE, were transformed in *E. coli* UA321 and presented exclusively on their own plasmids (pUA321:appDF and pUA626:SpoIIE) in order to examine the changes in regulation observed between them (Figure 2.1.). Mutation targets within the SpoIIE gene were selected by using the Qiagen CLC genomics workbench, which also aided in primer design. Two categorical types of mutations were made: large deletions and missense point mutations. Targets for genetic mutations were chosen based on their predicted effect on overall SpoIIE protein assembly, folding, and function. There are three post-translational interactions SpoIIE displays that served as central experimental targets: integration of monomers, dimerization, and activation of phosphatase. The restriction digest experiment removed the portion of the gene coding for the transmembrane anchor; thought to be imperative for integration. This is a stretch

of approximately 100 hydrophobic amino acids in the truncated version of SpoIIE present on the pUA626 plasmid. Site-direct mutants contained missense mutations leading to the translation of incorrect amino acids, potentially essential for dimerization or phosphatase activation. The missense point mutations were also chosen because of the change in fundamental properties exhibited by the amino acids. Two mutants were created, D274K and D260Q. DK is shorthand for a missense mutation changing aspartic acid to lysine at position 274 going from charged, acidic to charged, basic. DQ is a mutation changing glutamine to aspartic acid at 260, changing it from uncharged to charged, acidic. It is important to note that these position numbers correlate to the truncated version of SpoIIE present on the pUA626 plasmid, and not the full SpoIIE found in databases.

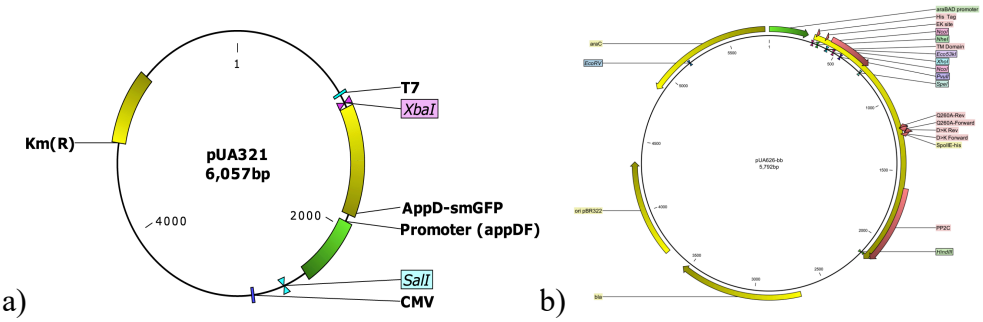

Figure 2.1. Plasmid maps with (a) appDF and(b) SpoIIE genes

As a proxy measurement for expression of the ABC peptide transporter in the presence of mutant SpoIIE, the plasmid with app has an upstream construct containing the reporter gene, GFP (green fluorescence protein). When expressed under normal conditions, in the absence of SpoIIE, this construct brightly fluoresced as appDF was promoted. In the positive control, where SpoIIE was unmutated, appDF was repressed along with the upstream construct, therefore showing little to no fluorescence. A discerning range of fluorescence levels were expected to be observed in this assay.

II.2. Materials and Methods

II.2.1. General culturing methods

Cultures of *E. coli* were grown in Luria Bertani (LB) growth media in the presence of carbenicillin (Cb), as the plasmids used carry the *bla* gene that makes them resistant; allowing for proper selection. Liquid media was made at 500mL per batch and carbenicillin was added post sterilization via autoclave, at a concentration of 50 ug/ml, denoted as Cb_{50}. Initial stocks of cell cultures and plasmids were grown in 25 mL of media at 37°C overnight in a shaking incubator, with the incubation time ranging from 12-16 hours. Glass flasks with ventilated caps were used to allowed for proper aeration. These liquid cultures were then used to either create glycerol stocks to be kept at -80°C, or processed via mini plus™ plasmid DNA extraction system (Viogene) in order to serve as control stocks of plasmid, kept at -20°C.

Plate culturing was performed on LB Cb plates, which were made in 500mL batches at a time, resulting in an average of 18 plates. Five-hundred microliters of Cb_{50} was added to the liquid media once cooling post autoclave sterilization. Plates were poured near open flame, allowed to set, and then stored at 4°C until used. Cultures of *E. coli* were spread onto LB Cb plates by first pipetting 50-100 µL of cell culture onto the plate surface and then streaking with a sterile loop, followed by incubation at 37°C overnight (12-16 hours).

II.2.2. Plasmid DNA extraction

Plasmid DNA extraction was performed as directed by the manufacturer protocol for mini plus™ plasmid DNA extraction system (Viogene) and the "spin method" was followed. This was done from either frozen stock of cell culture preserved in glycerol (800µL cell culture in 200µL 80% glycerol) at -20°C or directly from cell culture post-incubation. After plasmid extraction, samples were evaluated via gel electrophoresis. Specific stocks were set aside as controls, not to

be mutated. The quality and quantity of plasmid DNA of the control stocks was evaluated with Qubit™ 2.0 (Invitrogen) spectrophotometry and noted. All plasmid stocks were kept at -20°C throughout the progression of the project.

II.2.3. Gel electrophoresis

Gel electrophoresis often served as a qualitative measurement throughout this work for various purposes. Five grams of 2% agarose gel with 4µL SYBR® Safe DNA Gel Stain (Invitrogen) was used and ran at 100V for an average of 25 minutes in 1X TAE buffer acquired from a community lab stock (Tris base 4.844 g/L, acetic acid 1.21 mL/L, EDTA 0.372 g/L). Additionally, 5 µL of DNA MW 1Kb ladder (VWR) used from an individual stock stored at RT (10 µL ladder, 20 µL 6X gel loading dye (NEB), 80 µL 1X TAE buffer). A control plasmid DNA sample was always present in the first well for comparison. 6X loading dye was used by preference, as loading the wells tended to be more successful with lesser volumes. Images of the gels were captured using Kodak Electrophoresis Documentation and Analysis System in combination with ULTRA-LUM, Electronic UV Transilluminator. The images were then evaluated through an EOS program paired with the Nikon camera on a shared laboratory desktop computer.

II.2.4. Polymerase chain reaction (PCR) site-directed mutagenesis

Multiple site-directed mutations were carried out through Polymerase Chain Reaction (PCR). Quantification via Qubit™ 2.0 (Invitrogen) spectrophotometry was necessary before beginning, as PCR protocols require known DNA concentrations (ng/µL). PCR site-directed mutagenesis was preformed via the NEB Q5 protocol, and the primers were designed via Qiagen CLC genomics workbench. Using the primers shown in Table 2.1., site-directed mutagenesis via PCR was employed to create missense point mutations within SpoIIE on plasmid pUA626

extracted from *E. coli*. Each PCR reaction was performed in 25 µL, comprised of 12.5 µL Q5 hotstart high fidelity 2x mastermix (NEB), 1.25 µL of each F/R primer at a concentration of 10 µM, 1 µL of 1-25 ng of the template DNA, along with nuclease-free water to bring each reaction to the 25 µL final volume. PCR amplification was done using Thermal cycler (MWG-Biotech) and the reaction conditions were denaturation at 98°C for 30s, 25 cycles for annealing at 98°C for 10s, 50-72°C for 10-30s, 72°C for 20-30s/Kb and extension at 72°C for 2 min. These conditions were taken from the NEB Q5 manufacturer's protocol.

Table 2. 1. Primers used for PCR site-directed mutagenesis

Mutant	Primers	Sequence (5'→ 3')	Source
DK	F	CTTATCTCAAAAGTTAGAAGGCG	Eurofins
	R	GATTTAATGGATTTTGAAATACTT	
DQ	F	AATAGCAAATGCGATAAGAAGTATTTCAAAATC	Eurofins
	R	AACTTTCTACTTTCTGAAAAC	

*Primers designed using Qiagen CLC genomics workbench

II.2.5. Endonuclease restriction digest mutagenesis

An additional method for mutagenesis was used for the aim of deleting the transmembrane anchor region of the SpoIIE gene. The endonucleases used for this experiment were Spe1 and Eco53KI, chosen with the aid of the Qiagen CLC genomics workbench. Two identical reactions were done in tandem, one to be used for gel electrophoresis evaluation. Ten microliter reaction volumes were used, consisting of 8µL of plasmid pUA626 with control SpoIIE, 1µL cutsmart® (NEB) 0.5µL of SPE1, 0.5µL of Eco53KI, incubated in the Thermal cycler (MWG-Biotech) 37°C for 1 hour. Following incubation, one of the reactions was run on a gel against control pUA626 with conditions listed above. If the gel showed successful digest, then the secondary reaction tube was treated with 1µL of VENT DNA polymerase (NEB) and 1µL dNTP from a personal stock (10µL of A, T, C, G and 60µL of DI H_2O) and incubated at 72°C for 30 min followed by 3 min at 80°C. The step both inactivated the endonucleases and created blunt ends on the mutant plasmid. The reaction tube was then cooled to RT before adding 1µL T4 DNA ligase (NEB) and 1µL polynuclear kinase (Promega), left to incubate overnight.

Table 2. 2. Enzymes used in restriction digest mutagenesis

Enzyme	Restriction Site	Source
Spe1	A'CTAGT **(5'→ 3')** TGATC'A **(3'→ 5')**	NEB
Eco53KI	GAG'CTC **(5'→ 3')** CTC'GAG **(3'→ 5')**	NEB
SpoIIE TM	F: CCAGAGCGATTATTTAATAAG R: CTCGAGCTCGGATCCCCATCG	Eurofins

*Primers designed using Qiagen CLC genomics workbench

II.2.6. Mutant plasmid chemical transformation

In both cases of mutagenesis (site-directed and restriction digestion) after confirmation of mutation success via gel electrophoresis, the new mutant plasmid was transformed into chemically competent *E. coli* cells. Initially, *E. coli* 10-Beta competent cells (NEB) were used but yielded low transformation efficiency at 1.5 x 10^6 cfu/µg. Accordingly, transformation experiments were then carried out with NEB C2987 chemically competent *E. coli* cells, resulting in a much higher efficiency. Transformations were performed following the Q5 site-directed mutagenesis protocol (NEB) for all mutants. After a 1-hour incubation of transformed, chemically competent *E. coli* cells at 37°C, two LB Cb plates per transformation were streaked. Each plate as a minimum (50µL) and maximum (100µL) acceptable amount as stated in the transformation protocol. The transformed *E. coli* cells were pipetted onto the plate near open flame, and then spread either with a sterile loop or L-spreader. These plate cultures were incubated at 37°C overnight. After receiving the new, highly efficient competent cells, it was found that 50µL yielded an ideal number of colonies, making that the standard amount used.

II.2.7. Colony selection and broth cultures

Following the incubation of the transformed chemically competent *E. coli* (NEB C2987) cells on LB Cb plates, colonies were selected to grow in liquid media for further processing. Approximately 6 colonies were picked with a sterile loop and individually inoculated into 4mL of LB Cb liquid media in glass test tubes. These test tube broth cultures were incubated for 12-16 hours overnight at 37°C in a shaking incubator. Following incubation, 2mL of the cell culture was processed using mini plus™ plasmid DNA extraction system (Viogene). These highly concentrated mutant plasmid stocks were either immediately utilized or stored at -20°C for subsequent experiments.

II.2.8. Creation of electrocompetent *E. coli* stocks

Commercially purchased chemically competent *E. coli* cells (NEB C2987) were used in mutant transformations, but the use of electroporation to create electrocompetent *E. coli* cells was necessary to observe the regulatory relationship in question. This is because the *E. coli* strain (pUA321) containing the plasmid with the app locus (peptide transport gene) is not store-bought, but rather developed by the Ivey lab. Therefore, it must be made electrocompetent via electroporation so that transformation of the plasmids containing the SpoIIE mutants (pUA626) yields a new *E. coli* UA321 strain where both plasmids are present (pUA321 and pUA626).

E. coli UA321 was grown in liquid culture by inoculating 25mL Km_{50} LB broth with 20µL of glycerol stock and then incubating for 12-16 hours at 37°C. Post-incubation, 1mL was used to inoculate 50mL LB broth with 50µL of Km_{50}. This liquid culture was permitted to grow into late log phase, or until a Klett value of 100-150 was reached (approx. 4 hours). At this point, the culture was kept on ice along with 50mL centrifuge tubes, 10% glycerol stock, and sterile DI H_2O. After multiple centrifugations (11,000rpm at 10 min) and rebalancing with sterile water, the pellet was suspended in 0.4mL of 10% glycerol, this then was divided up into approximately 6 tubes of 100µL glycerol stocks to be kept at -80°C until transformation via electroporation.

II.2.9. Mutant transformation to *E. coli* UA321 and fluorescence prep

In order to evaluate the regulatory relationship observed between SpoIIE and app, the plasmids with mutant SpoIIE were then transformed into *E. coli* strain UA321, which contained the app locus on its own plasmid, pUA321. This transformation was carried out via electroporation. The electrocompetent *E. coli* cells were kept on ice, alongside other materials such as the cuvettes and cuvette holder. One-hundred microliters of electrocompetent E. coli UA321 plus 1µL of mutant plasmid pUA626 were shocked using the following settings: 1.5 kV,

200 ohms resistance, and 25microfarads capacitance, using Bio-Rad Gene Pulser™ and then allowed to incubate for an hour at 37°C in 1000mL of recovery media.

After transformation, 50-100 µL cell culture was pipetted onto prewarmed Cb/ Km LB plates and then spread with a sterile L-spreader near open flame. 25 mL of Km_{25} was added to the Cb LB plates previously described, by spreading 10µL Km onto the plates' surface 5 minutes before the introduction of the transformed electrocompetent *E. coli* UA321. Both antibiotics are used in this transformation because the plasmid with the SpoIIE mutant is resistant to Cb, while the plasmid containing the app locus is resistant to Km; therefore colonies present after incubation will contain both plasmids, as desired. The Cb/ Km LB plates were incubated at 37°C for 12-16 hours overnight.

The next day, 4-6 colonies were selected to subculture and spread onto a subsequent Cb/ Km LB plate, and their size was noted. The new plate was sectioned off into six partitions and each selected colony was spread into a designated section, two of which were used for controls.

The two controls (dim and bright) were streaked from freezer stocks previously cultured and transformed through methods mentioned above. The "bright" control was *E. coli* UA321 in absence of SpoIIE: no repression. While the "dim" control was *E. coli* UA321 in the presence of unmutated SpoIIE on pUA626: strong repression. The freezer stocks were thawed on ice and then a loopful was to be streaked onto the plate. After all picked colonies were streaked, the plate was to be incubated at 37°C for 12-16 hours before observing fluorescence.

II.2.1. Evaluation of fluorescence

Fluorescence was detected and recorded by using the Canon EOS Utility on a shared lab desktop Dell connected to a Cannon Rebel T4 camera. Blue light provided by 420 nm wavelength LED light strips was used for fluorescence excitatation. The sectioned plate, as

describe in the section above, was removed from incubation and placed with its lid off inside the viewing arena within the black box. The relative brightness was to be scored against the two controls. Note: fluorescence methods were only practiced with, no useable fluorescence data was collected.

II.3. Results

II.3.1. Creation of *C. difficile* SpoIIE mutants

After mutagenesis, the mutants DK and DQ were first evaluated via gel electrophoresis before further transformations were carried out. The mutated plasmid DNA was stored at -20°C with the intention of sending them for sequencing in order to confirm that they were successfully mutated in the manner intended. The endonuclease restriction digest mutant aimed at deleting the transmembrane anchor region proved to be trickier than the missense point mutations. The evaluation of this mutant via gel electrophoresis was expected to have yielded obvious results with such a large sequence missing. Many tactics were employed to try to get this restriction digest to work. Initially, new endonucleases were ordered in case the ones used from the freezer stocks were inactive. Second, the protocol outlined in the methods section was tinkered with numerous times, extending incubations or increasing various enzyme volumes. Third, specific PCR primers (instead of the endonucleases) were designed using the Qiagen CLC Genomics Workbench. These primers have not yet been tested because of the lab closures.

II.3.2. Experimental progress

Unfortunately, this portion of the project came to an abrupt halt and as a result, no concrete data was gathered for analysis of the mutants. In the early Spring of 2020, the University of Arkansas labs closed for non-essential research due to the on-going Covid-19 pandemic which impacted many research projects, including this one. In light of the pandemic, a second research aim, and therefore third chapter, was pursued in order to supplement the loses exhibited in this section. The additional research aim was geared to be achieved in the absence of a laboratory and focused on computational modeling of the protein structures resultant of the genetic mutations outlined in this chapter.

II.4. Discussion

As mentioned, the intentions of this research aim were to assess protein function following mutagenesis experimentation. The presence of GFP on the plasmid containing appDF, fundamentally provided a qualitative measurement of expression for appDF in the presence of mutated SpoIIE. The level of repression would parallel to the relative level of fluorescence produced by the *E. coli* colonies transformed with the modified plasmids. Having two controls, the "dim" and "bright" control, would have allowed for the comparison of the brightness exhibited by the mutants when streaked and grown upon the partitioned plate described. The experimental design for the creation of the *C. difficile* SpoIIE mutants required the exercise of many classical microbiology and molecular biology techniques, but the lack of data collected proved to be disappointing.

II.5. Conclusion

The in vitro goals pursued in this research project ultimately lack conclusive evidence to either support or reject the stated hypothesis. The absence of results is not due to experimental failure, but rather external and circumstantial reasons. This leaves room for incoming graduate researchers to confidently complete this mutagenesis assay, which will be discussed in Chapter IV: Future Directions.

III. COMPUTATIONAL MODELING OF SPOIIE STRUCTURE

III.1. Introduction

A growing number of biological sciences are beginning to fundamentally incorporate and rely on in silico techniques to substantiate in vitro results. The structural modeling of proteins is an example of how computational techniques provide unique visual representations of molecular interactions otherwise limitedly studied in the lab. Three-dimensional images of proteins can aid in the visualization of complex interactions. These images show secondary protein structures such as helices or sheets, and their resulting tertiary arrangement (Bienert et al., 2017; Martí-Renom et al., 2000). Being able to conceptualize these structures elucidates the mechanisms underlying enzymatic events such as active site binding and protein-protein interactions.

Modeling of structurally unknown or "target" proteins is done by taking the DNA sequence data or the primary amino acid sequence and aligning it with the sequence of a structurally similar template (Martí-Renom et al., 2000). This predictive technique is known as homology or comparative modeling. The quality and accuracy of the structures generated through homology modeling is dependent upon the sequence identity between the target and the template. Alignments with low E-values signify that the template and target protein are closely evolutionarily related (Martí-Renom et al., 2000). Protein tertiary structures are more conserved than DNA sequences, and a sequence identity above 20% is ideal and renders significantly similar protein structures that share similar properties (Martí-Renom et al., 2000; Chothia & Lesk 1986). Inaccuracies in homology modeling are derived from initial errors in template selection and alignment, so the choice of template is a crucial step.

In this study, SpoIIE from the analogous model organism, *Bacillus subtilis* is used as the template, as this protein structure is known. The phosphatase domain (C' terminus) of SpoIIE has previously been crystallized and the structure is readily available in online protein databases (PDB) to use as the template. After the production of the target *C. difficile* SpoIIE phosphatase domain, models of the unknown central domain and N'-terminus TM domain were created through non-template threading, which is a method that produces models based off structural homologs in the PDB, rather than relying purely on sequence similarity. This created a novel structure comprised of all three protein domains.

Prior to the creation of this complete SpoIIE model, multiple algorithms were assessed and compared for their initial sequence alignment and the resulting models. This was done by comparing sequence identity, model quality estimates, clash values, and then the side-by-side visualization of the model variants. Additionally, after the webservices were analyzed, preferred ones were chosen to model the target *C. difficile* SpoIIE mutants. Qualitative conclusions based upon the structural models are discussed in this section.

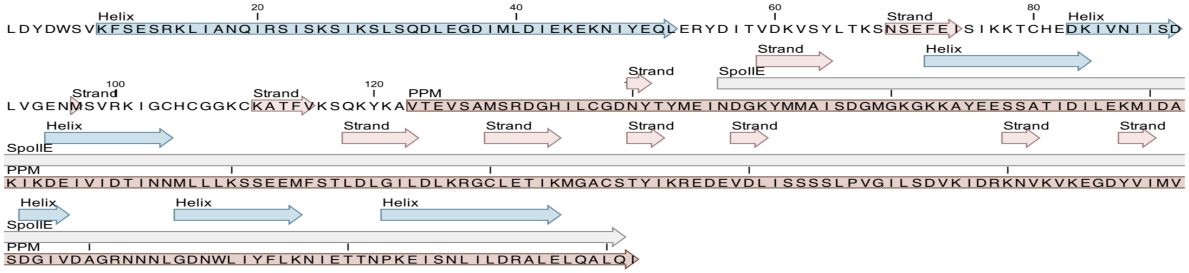

Figure 3.1. Primary sequence: *C. difficile* SpoIIE, UA001

***Image created using Qiagen CLC Genomic Workbench**

III.2. Materials and Methods

III.2.1. Sequence alignment and template selection

The first step in homology modeling is to find a suitable template from which the predictive model is based. The template was predicted to be that of *B. subtilis* SpoIIE, as it is comparatively used in countless studies paralleled to *C. difficile*. Despite this projection, the primary sequence, or FASTA file, of the lab's truncated SpoIIE was BLAST against multiple databases in order to confirm a reliable template was used for homology modeling. This step was done in combination with modeling as the web servers used are full pipelines, preforming all required steps to produce a model. The sequence identities derived from these searches are reported in Results. As an additional independent search, a BLAST was run against the EMBL (European Molecular Biology Laboratory) database, which yielded similar top results of *B. subtilis* SpoIIE strain 168 with a sequence identity of 27.5% and an E-value of 3.5e-50. Sequence identities must be above 20% in order to use homology modeling. The sequence coverages are almost entirely for the phosphatase domain, whose crystal structure exists in the PDB. Across multiple services, this domain's coverage was upwards of 95%.

III.2.2. Modeling programs

Multiple bioinformatics software services, accessed through free online servers, were used to develop structural models of *C. difficile* SpoIIE through homology modeling against a *B. subtilis* template. Each of the websites used have their own algorithms, which allowed for comparison of scores and model constraints. The websites used were Swissmodel, Phyre and Robetta. After the modeling softwares were analyzed for their respective capabilities, the structural mutants were then modeled and compared.

III.3. Results

III.3.1. Homology modeling

The FASTA sequences of *C. difficile* SpoIIE (Appendix Figure A 1.) and *B. subtilis* (Appendix Figure A 2.) were aligned. The sequence for *B. subtilis* is notably longer, because as mentioned in Chapter 2, the *C. difficile* SpoIIE sequence from pUA626 is truncated and missing a little over 200 amino acids from the complete TM sequence. The *C. difficile* SpoIIE FASTA consists of multiple domains, of which the C'-terminus or phosphatase domain is the most studied. In order from N'→ C' the sequence shows ~4 TM helices, a linker region, an alpha-helix, and the regulatory and catalytic phosphatase domains (PPM). The sequence -LDYD- in the *C. difficile* SpoIIE FASTA signifies the beginning of the phosphatase domain. When preforming homology modeling, the FASTA sequence was cut after -LDYD- at amino acid number 245, in order to report the highest coverage and confidence level. This begins the phosphatase domain, which has previously been crystallized and the X-ray crystallization structure data is available in these databases to use as the template for the *C. difficile* SpoIIE being studied in this project. The reported coverages after cutting the FASTA at amino acid 245, were all over 95%, which is ideal for homology modeling.

Three bioinformatic programs were used: Swissmodel, Robetta, and Phyre. These are all full pipeline webservices, meaning that they not only find a suitable template, but also execute a pairwise sequence alignment with the target, and produce the resultant homology model. These algorithms also report model evaluation values assessing their quality and constraints, which were considered before continuing on with mutation modeling.

The homology models from Swissmodel, Robetta and Phyre of the phosphatase domain appear nearly identical, which is a nice visual validation for algorithm accuracy. These images are shown side-by-side in Figure 3.2. It is important to note that the model produced by Phyre is not dimerized like the other two images. Numerically, the models prove to be similar as well, with identical sequence similarity, coverage, and confidence levels. The homology modeling of the SpoIIE phosophatase domain is reliable and was preformed with ease due to the presence of the *B. subtilis* homolog in the PDB.

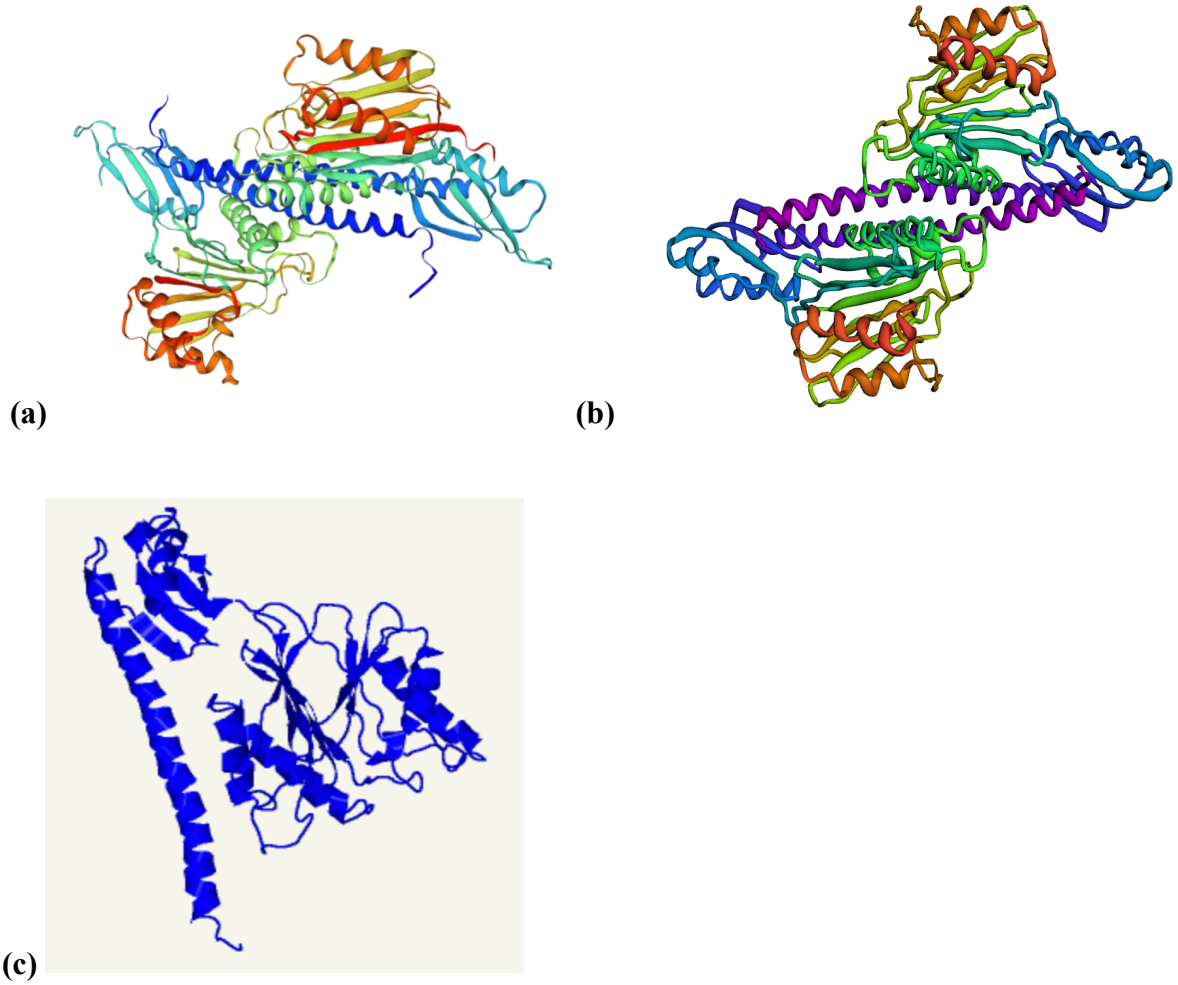

Figure 3.2. *C. difficile* SpoIIE phosphatase domain: Swissmodel (a), Robetta (b), Phyre (c)

After modeling the phosphatase domain, the N'-terminus or truncated TM region (first 100 amino acids) was also homology modeled through Phyre (Figure 3.3.) The generated image looks exactly as expected, with four TM helices. Interestingly, the template used for this target sequence was a thiamine binding protein with 25% sequence similarity. Phyre modeled 72% (coverage) of the TM region FASTA sequence with 94.3% confidence. Phyre was chosen as the program for this homology model as the job was completed the fastest, and created the best visual representation.

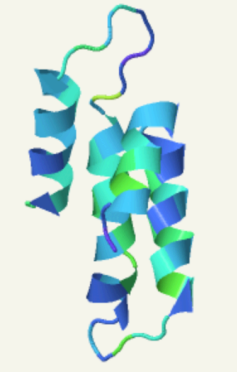

Figure 3.3. *C. difficile* SpoIIE transmembrane

After creating the initial homology models, the phosphatase domain was further evaluated using Phyre, and the analyses are shown in Figure 3.4. Phyre[2] and SuSPect are investigative sister programs that have the ability to assess point mutation effect on structure (a), the potential for steric clashes (b), as well as provide a Ramachandran analysis (c). The coloring scheme of the models below show red as indicating the most problematic points.

The reports generated and shown in Figure 3.4 reflect the missense mutations outlined in Chapter 2. Image (d) is of amino acid 260, specifically looking at mutant DQ, while image (e) specifically shows amino acid 274, or mutant DK. These bar graphs show which point mutations are likely to be the most deleterious at the indicated point in the FASTA sequence. By comparison, Phyre is predicting the DQ point mutation to be more likely to affect protein function than the DK mutation.

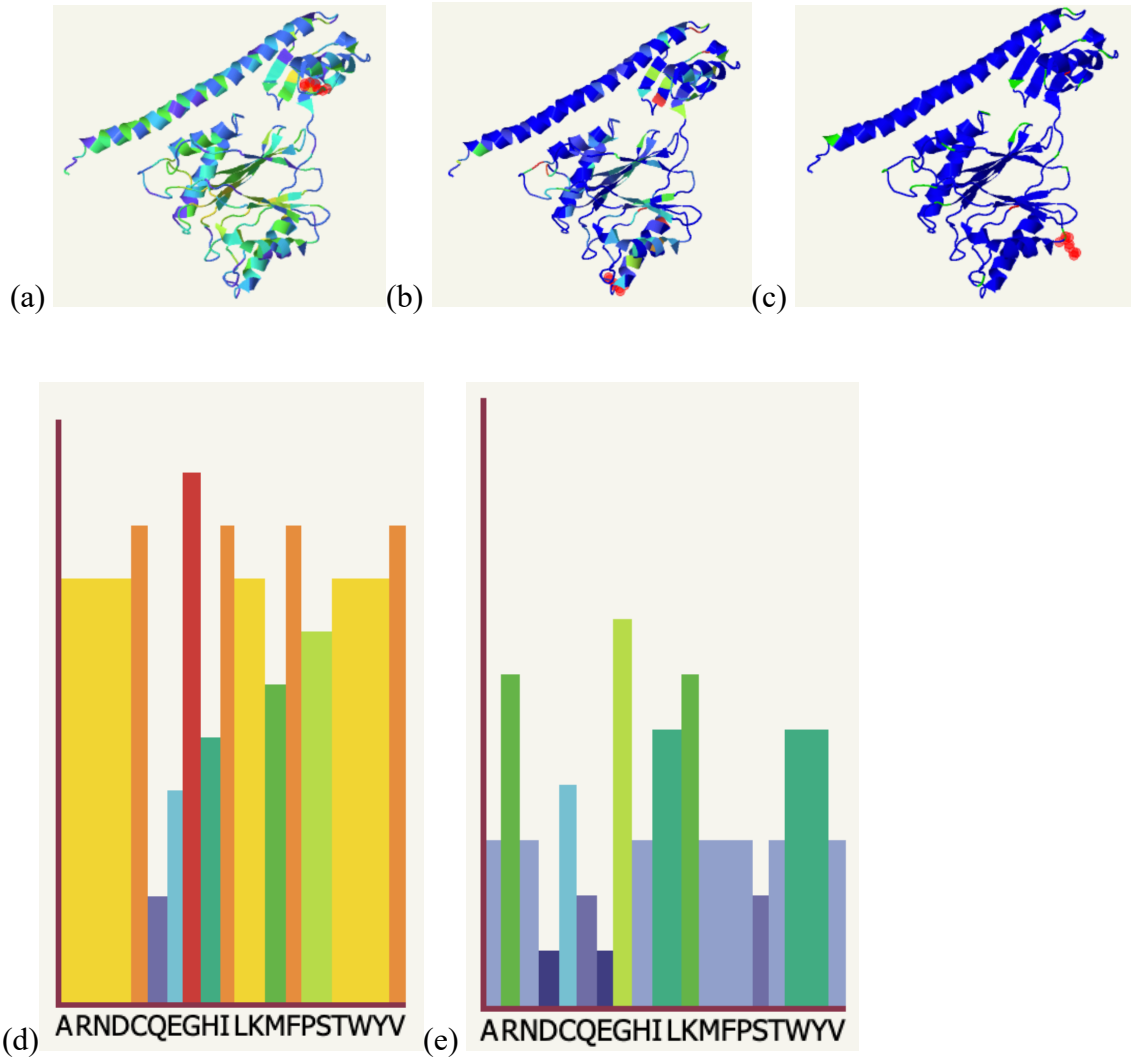

Figure 3.4. *C. difficile* SpoIIE phosphatase domain analysis: Phyre[2]/SuSPect

III.3.2. Mutant SpoIIE structural models

After the primary homology modeling was done through multiple servers, mutant models were then created based upon the point mutations outlined in Chapter 2. The mutant models were created using Phyre, because of the additional capabilities of Phyre[2] and SuSPect, as previously mentioned. Mutants models were made by manually changing the FASTA sequence before running the program. As shown in Figure 3.5. and 3.6. the structural conformation of the SpoIIE phosphatase domain does not appear to change because of the point mutations. As an interesting way to double check these mutants, the residue preference at these specific points was analyzed and showed high preference towards the original amino acid. This bar graph is reported next to the corresponding mutant model.

The TM mutant was not modeled because it cannot be evaluated for loss of function in the same manner as the point mutations. The TM mutation is very different in nature and should fundamentally be evaluated in vitro.

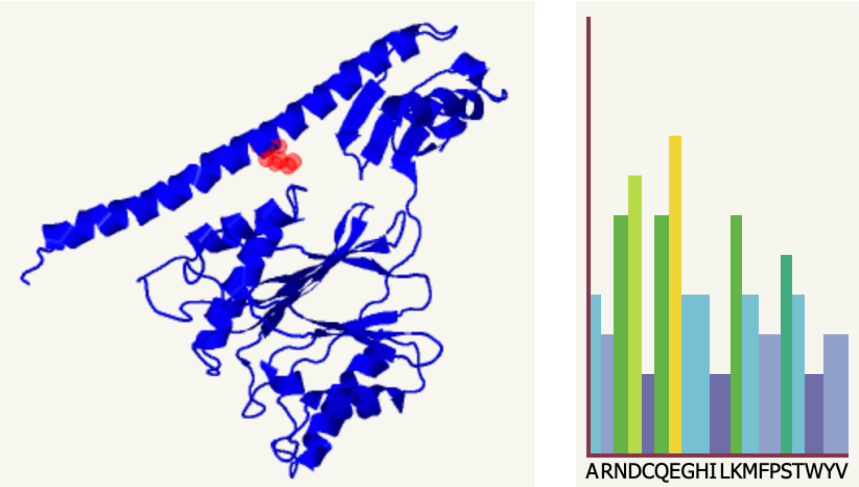

Figure 3.5. *C. difficile* SpoIIE DK mutant model: Phyre[2] and SuSPect

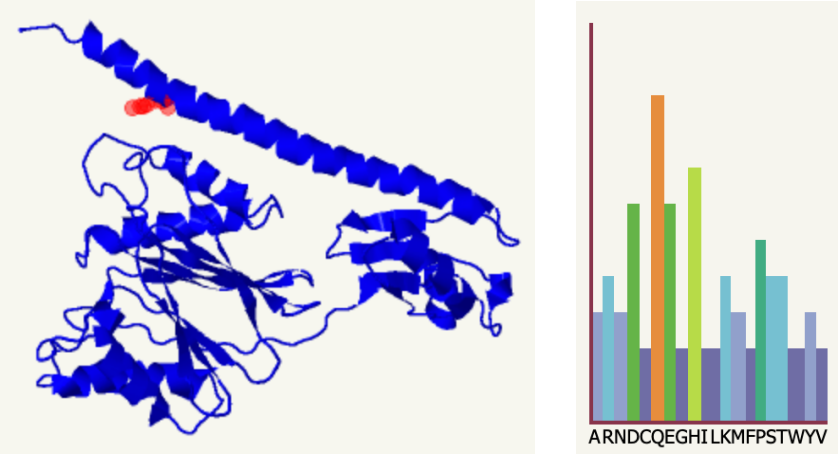

Figure 3.6. *C. difficile* SpoIIE DQ mutant model: Phyre[2] and SuSPect

Importantly, Phyre only shows SpoIIE as a monomer and the point mutations being analyzed are on the alpha-helix. This alpha-helix is known to serve as the dimer interface, so questions were raised as to whether or not these point mutations could be affecting dimerization. This is a functional flaw of the Phyre program, as the mutant models appeared to not have conformational changes existing as a monomer, but could very well show differences after forming a dimer.

III.3.3. Non-template threading

In addition to homology modeling, the webservice, Robetta, has a higher predictive capability in the sense that it can construct non-template models through threading and fold-recognition programming. Robetta was able to predict the structure of the full FASTA sequence, and not just phosphatase domain. Threading programs have the unique ability to produce models with or without sequence similarity; template-free. These types of alorithms search the PDB for structural homologs and folding pattern similarities rather than just relying on sequence similarity.

This created a novel model of *C. difficile* SpoIIE, spanning the entire truncated lab version of SpoIIE. There is no comparative structural protein data on the SpoIIE central linker/ switch domain via NMR or X-ray crystallization, so the algorithm made its modeling predictions based off the target FASTA sequence and structural homologs in the PDB. This method is structurally not the most reliable, but the generated model is shown in Figure 3.9. The reported confidence level is 57%. The TM region is depicted in purple, the linker/ central domain in blue, and the phosphatase domain spans green through red. The phosphatase domain is recognizably similar in this model to the other homology models shown in this chapter. It can be assumed that a large portion of the 57% confidence level is due to the data available on the phosphatase domain, which accounts for over half of the structural model.

Interestingly, the TM membranes modeled by this program do not show the four TM helices as one would predict these hydrophobic regions to appear. Robetta shows the truncated TM SpoIIE gene fragment to possibly be encoding a soluble, dimeric form of SpoIIE. This is highly speculative, but if this model is correct then the soluble form of SpoIIE could be isolated and crystallized. This hopeful find for the lab is discussed in Future Directions.

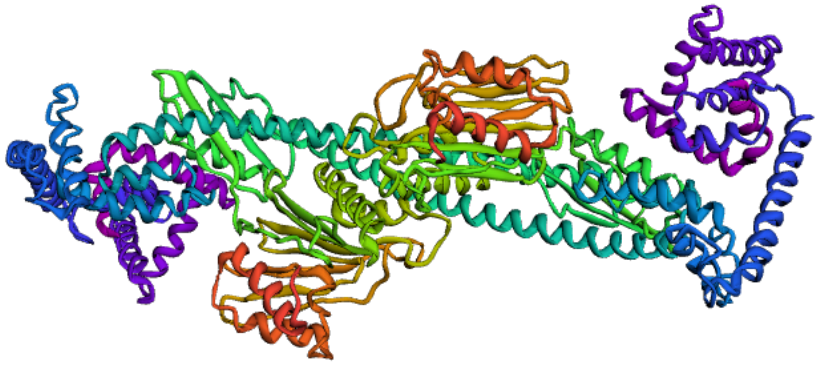

Figure 3.9. *C. difficile* SpoIIE complete model: Robetta

III.4. Discussion

Protein structure determination through experimental methods such as NMR (nuclear magnetic resonance) or X-ray crystallography is time-consuming and costly, but newer, progressive techniques have taken main stage as computational modeling further develops. Knowledge based methods like homology modeling are increasing popular and provide an efficient alternative to determining protein structures.

The modeling of *C. difficile* SpoIIE exhibited in this study confidently depicts the phosphatase domain through homology modeling with a template structure derived from *B. subtilis* SpoIIE. Across all webservices used for homology modeling, the confidence level was above 95%. The TM region was also accurately modeled, portraying four transmembrane helices, as expected. The SpoIIE gene used in these experiments was known to be the truncated version, having just four of the ten TM helices. Comparatively, Robetta modeled this same TM region quite differently, showing a cluster of helices in a formation not observed in highly hydrophobic domains.

In order to model the central linker domain of *C. difficile* SpoIIE, a different approach had to be taken as there is no template data available for homology modeling. The program Robetta was used to model this region template-free through protein threading. This developed a complete predicted image of SpoIIE. This model was made by an algorithm that searches the PDB for folding and structural homologs in order to develop a comparative model for the portions of the target sequence lacking a homologous template sequence.

After SpoIIE was modeled through various methods, the point mutations outlined in Chapter 2 were evaluated by Phyre. According to Phyre, the point mutation denoted as DQ is likely to have a deleterious effect on protein function. Excitingly, this can be tested for appDF repression in vitro upon lab reopening by completing the fluorescence assay. Contrastingly, the DK mutant is not projected to effect function to the same extent. These predictions are an exciting way to visualize the effects of the PCR site-directed mutagenesis experiments discussed in Chapter 2. The relative fluorescence levels in vitro may mirror this prediction, where for example, the lesser effect on function in the DK mutant may still fluoresce at a level similar to the dim control as this SpoIIE mutant may still functionally repress appDF.

Alternatively, Phyre was unable to predict conformational changes that may occur once the monomeric form of SpoIIE is dimerized. This disadvantage can be overcome by further structural investigation by other mutant modeling programs, and is an example of how the preferred modeling program to use ultimately depends on the purpose and goal of the project.

As shown by modeling the mutations done in vitro, each program used and their respective algorithms have their advantages and disadvantages. Modeling of protein structure is a useful method for investigating unknown relationships, such as the repression of appDF by SpoIIE, as understanding structure is often the first step to understanding function.

III.5. Conclusion

The data presented through computational modeling of mutant SpoIIE elucidates the experimental mutagenesis effect on protein function. The greatest effect on function should be seen in the DQ mutant, where aspartic acid is translated rather than glutamine. Comparatively, the DK mutant is projected to have a lesser effect on SpoIIE protein function. The mutation analysis program that was used in this study produced a spectrum of predictive effects. This has the potential to mirror the spectrum of relative brightness as would be seen in the fluorescence assay. It could be that the DK mutant may fluoresce to a level brighter than the DQ mutant, but dimmer than the "bright" control. The computational analyses done in this project could effectively complement and substantiate fluorescence data.

It must be noted that the predicted mutational effects simply project whether a mutation has a statistical probability of interfering with protein function, but do not have the capacity to forecast *which* function. This leaves ambiguity as SpoIIE is a multifunctional protein. If the fluorescence assay results do not match the computational predictions, then this project could provide insight into SpoIIE function beyond appDF repression or involvement in peptide transport. Both point mutations were located on the cytoplasmic alpha-helix which is known to be involved in the timing and regulation of the phosphatase domain activity. This long alpha helix serves as the dimer interface, so if the DK or DQ mutations happen to interfere with dimerization, then an entire cascade of sporulation events may be inhibited.

Regardless of the outcome for future experimentation, the protein modeling of *C. difficile* SpoIIE has proven to be a useful tool and could aid in furthering the holistic understanding for the pathogenicity and virulence factors of this infectious disease.

IV. FUTURE DIRECTIONS

In light of the results presented from both of the stated research aims, future investigation into this topic may take three presumable routes. The first is quite simple, in that incoming graduate researchers could finish the mutagenesis fluorescence assay. The techniques used are known to be successful by previous students and all materials are abundantly present in the lab. In addition to evaluating the mutants from this project, future students could also create new ones. The qualitative fluorescence data to be gathered, may also be interpreted in light of the structural models created and discussed in this research project. It is expected that application of these visual structures may bolster future lab projects reliant on in vitro experiments.

Second, it should be investigated as to how these proteins are interacting specifically within *E. coli*. SpoIIE is a multifunctional protein, and it should be assumed that it is dynamically interacting with many other proteins. This scenario must always be considered when not working in the native cell, here being *C. difficile*. For instance, it is known that the SpoIIE C-terminal cytoplasmic domains carry out specific binding to cytoskeletal proteins such as FtsZ, RodZ, and others (Ben-Yehuda & Losick 2002; Lucet et al., 2000). In the analogous system *Bacillus subtilis*, SpoIIE interacts with FtsZ and Rodz to aid in spore septum formation, but these interactions are far from understood (Arigoni et al., 1999; Ben-Yehuda & Losick 2002; Lucet et al., 2000). FtsZ is present in *E. coli*, so it can be inferred that SpoIIE is involved in multiple regulatory relationships, not just the one with app focused on in this study. The structural models of the SpoIIE protein presented in this project, could be applied to further proteomic studies ideally investigating how transformation to competent *E. coli* affects this novel relationship between SpoIIE and app, in the presence of other proteins such as FtsZ.

Lastly, if the model produced by Robetta is correct, then the SpoIIE gene fragment used experimentally in the Ivey lab may be encoding a soluble form of the SpoIIE protein, which is normally membrane bound by the hydrophobic N'-terminus transmembrane region. This freely dispersed, soluble form of SpoIIE could then be isolated, purified, and crystallized. This is an exciting possible future direction for the lab as the crystal structure for *C. difficile* SpoIIE is presently unknown. Ultimately, the work presented in this thesis may contribute to may future research projects and potential directions for the Ivey lab.

V. REFERENCES

Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. Introduction to Pathogens. Available from: https://www.ncbi.nlm.nih.gov/books/NBK26917/

Antonovics, J., Wilson, A. J., Forbes, M. R., Hauffe, H. C., Kallio, E. R., Leggett, H. C., Longdon, B., Okamura, B., Sait, S. M., & Webster, J. P. (2017). The evolution of transmission mode. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *372*(1719), 20160083. https://doi.org/10.1098/rstb.2016.0083

Arigoni, F., Guérout-Fleury, A. M., Barák, I., & Stragier, P. (1999). The SpoIIE phosphatase, the sporulation septum and the establishment of forespore-specific transcription in Bacillus subtilis: a reassessment. *Molecular microbiology*, *31*(5), 1407–1415. https://doi.org/10.1046/j.1365-2958.1999.01282.x

Balloux, F., & van Dorp, L. (2017). Q&A: What are pathogens, and what have they done to and for us? *BMC biology*, *15*(1), 91. https://doi.org/10.1186/s12915-017-0433-z

Barák, I., Behari, J., Olmedo, G., Guzmán, P., Brown, D. P., Castro, E., Walker, D., Westpheling, J., & Youngman, P. (1996). Structure and function of the Bacillus SpoIIE protein and its localization to sites of sporulation septum assembly. *Molecular microbiology*, *19*(5), 1047–1060. https://doi.org/10.1046/j.1365-2958.1996.433963.x

Barreto, M. L., Teixeira, M. G., & Carmo, E. H. (2006). Infectious diseases epidemiology. *Journal of epidemiology and community health*, *60*(3), 192–195. https://doi.org/10.1136/jech.2003.011593

Bartlett, J.G. & Gerding, D. N. Clinical Recognition and Diagnosis of *C. difficile* Infection, *Clinical Infectious Diseases*, Volume 46, Issue Supplement_1, January 2008, Pages S12–S18, https://doi.org/10.1086/521863

Beis K. (2015). Structural basis for the mechanism of ABC transporters. *Biochemical Society transactions*, *43*(5), 889–893. https://doi.org/10.1042/BST20150047

Ben-Yehuda, S., & Losick, R. (2002). Asymmetric cell division in B. subtilis involves a spiral-like intermediate of the cytokinetic protein FtsZ. *Cell*, *109*(2), 257–266. https://doi.org/10.1016/s0092-8674(02)00698-0

Bienert, S., Waterhouse, A., de Beer, T. A., Tauriello, G., Studer, G., Bordoli, L., & Schwede, T. (2017). The SWISS-MODEL Repository-new features and functionality. *Nucleic acids research*, *45*(D1), D313–D319. https://doi.org/10.1093/nar/gkw1132

Borriello, S. P., Davies, H. A., Kamiya, S., Reed, P. J., & Seddon, S. (1990). Virulence factors of C. difficile. *Reviews of infectious diseases*, *12 Suppl 2*, S185–S191. https://doi.org/10.1093/clinids/12.supplement_2.s185

Bouza E. (2012). Consequences of C. difficile infection: understanding the healthcare burden. Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 18 Suppl 6, 5–12. https://doi.org/10.1111/1469-0691.12064

Brown, S. P., Cornforth, D. M., & Mideo, N. (2012). Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. *Trends in microbiology*, *20*(7), 336–342. https://doi.org/10.1016/j.tim.2012.04.005

Carman, R. J., Stevens, A. L., Lyerly, M. W., Hiltonsmith, M. F., Stiles, B. G., & Wilkins, T. D. (2011). C. difficile binary toxin (CDT) and diarrhea. *Anaerobe*, *17*(4), 161–165. https://doi.org/10.1016/j.anaerobe.2011.02.005

Carniol, K., Ben-Yehuda, S., King, N., & Losick, R. (2005). Genetic dissection of the sporulation protein SpoIIE and its role in asymmetric division in Bacillus subtilis. *Journal of bacteriology*, *187*(10), 3511–3520. https://doi.org/10.1128/JB.187.10.3511-3520.2005

Casadevall, A., & Pirofski, L. A. (2000). Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. *Infection and immunity*, *68*(12), 6511–6518. https://doi.org/10.1128/iai.68.12.6511-6518.2000

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography*, *66*(Pt 1), 12–21. https://doi.org/10.1107/S0907444909042073

Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO journal*, *5*(4), 823–826.

Cyprowski, M., Stobnicka-Kupiec, A., Ławniczek-Wałczyk, A., Bakal-Kijek, A., Gołofit-Szymczak, M., & Górny, R. L. (2018). Anaerobic bacteria in wastewater treatment plant. *International archives of occupational and environmental health*, *91*(5), 571–579. https://doi.org/10.1007/s00420-018-1307-6

Deakin, L. J., Clare, S., Fagan, R. P., Dawson, L. F., Pickard, D. J., West, M. R., Wren, B. W., Fairweather, N. F., Dougan, G., & Lawley, T. D. (2012). The C. difficile spo0A gene is a persistence and transmission factor. *Infection and immunity*, *80*(8), 2704–2711. https://doi.org/10.1128/IAI.00147-12

Denève, C., Janoir, C., Poilane, I., Fantinato, C., & Collignon, A. (2009). New trends in C. difficile virulence and pathogenesis. *International journal of antimicrobial agents*, *33 Suppl 1*, S24–S28. https://doi.org/10.1016/S0924-8579(09)70012-3

Depestel, D. D., & Aronoff, D. M. (2013). Epidemiology of C. difficile infection. *Journal of pharmacy practice*, *26*(5), 464–475. https://doi.org/10.1177/0897190013499521

Diard, M., & Hardt, W. D. (2017). Basic Processes in *Salmonella*-Host Interactions: Within-Host Evolution and the Transmission of the Virulent Genotype. *Microbiology spectrum*, *5*(5), 10.1128/microbiolspec.MTBP-0012-2016. https://doi.org/10.1128/microbiolspec.MTBP-0012-2016

Di Bella, S., Ascenzi, P., Siarakas, S., Petrosillo, N., & di Masi, A. (2016). C. difficile Toxins A and B: Insights into Pathogenic Properties and Extraintestinal Effects. *Toxins*, *8*(5), 134. https://doi.org/10.3390/toxins8050134

Dodson, A. P., & Borriello, S. P. (1996). C. difficile infection of the gut. *Journal of clinical pathology*, *49*(7), 529–532. https://doi.org/10.1136/jcp.49.7.529

Dubberke, E. R., & Olsen, M. A. (2012). Burden of C. difficile on the healthcare system. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, *55 Suppl 2*(Suppl 2), S88–S92. https://doi.org/10.1093/cid/cis335

Dunn, R. R., Davies, T. J., Harris, N. C., & Gavin, M. C. (2010). Global drivers of human pathogen richness and prevalence. *Proceedings. Biological sciences*, *277*(1694), 2587–2595. https://doi.org/10.1098/rspb.2010.0340

Edwards, A. N., & McBride, S. M. (2014). Initiation of sporulation in C. difficile: a twist on the classic model. *FEMS microbiology letters*, *358*(2), 110–118. https://doi.org/10.1111/1574-6968.12499

Edwards, A. N., Nawrocki, K. L., & McBride, S. M. (2014). Conserved oligopeptide permeases modulate sporulation initiation in C. difficile. *Infection and immunity*, *82*(10), 4276–4291. https://doi.org/10.1128/IAI.02323-14

Engering, A., Hogerwerf, L., & Slingenbergh, J. (2013). Pathogen-host-environment interplay and disease emergence. *Emerging microbes & infections*, *2*(2), e5. https://doi.org/10.1038/emi.2013.5

Fawley, W. N., Underwood, S., Freeman, J., Baines, S. D., Saxton, K., Stephenson, K., Owens, R. C., Jr, & Wilcox, M. H. (2007). Efficacy of hospital cleaning agents and germicides against epidemic C. difficile strains. *Infection control and hospital epidemiology*, *28*(8), 920–925. https://doi.org/10.1086/519201

Fimlaid KA, Bond JP, Schutz KC, Putnam EE, Leung JM, Lawley TD, et al. (2013) Global Analysis of the Sporulation Pathway of *C. difficile*. PLoS Genet 9(8): e1003660. https://doi.org/10.1371/journal.pgen.1003660

Fimlaid KA, Jensen O, Donnelly ML, Siegrist MS, Shen A (2015) Regulation of *C. difficile* Spore Formation by the SpoIIQ and SpoIIIA Proteins. PLoS Genet 11(10): e1005562. https://doi.org/10.1371/journal.pgen.1005562

Freeman, J., Bauer, M. P., Baines, S. D., Corver, J., Fawley, W. N., Goorhuis, B., Kuijper, E. J., & Wilcox, M. H. (2010). The changing epidemiology of C. difficile infections. *Clinical microbiology reviews*, *23*(3), 529–549. https://doi.org/10.1128/CMR.00082-09

Gao, X., Wang, H., Wang, H., Qin, H., & Xiao, J. (2016). Land use and soil contamination with Toxoplasma gondii oocysts in urban areas. *The Science of the total environment*, *568*, 1086–1091. https://doi.org/10.1016/j.scitotenv.2016.06.165

Green, S. K., Schroth, M. N., Cho, J. J., Kominos, S. K., & Vitanza-jack, V. B. (1974). Agricultural plants and soil as a reservoir for Pseudomonas aeruginosa. *Applied microbiology*, *28*(6), 987–991.

Hassoun, A., Linden, P. K., & Friedman, B. (2017). Incidence, prevalence, and management of MRSA bacteremia across patient populations-a review of recent developments in MRSA management and treatment. *Critical care (London, England)*, *21*(1), 211. https://doi.org/10.1186/s13054-017-1801-3

Hecht, G., Pothoulakis, C., LaMont, J. T., & Madara, J. L. (1988). C. difficile toxin A perturbs cytoskeletal structure and tight junction permeability of cultured human intestinal epithelial monolayers. *The Journal of clinical investigation*, *82*(5), 1516–1524. https://doi.org/10.1172/JCI113760

Jacobs, S. E., Lamson, D. M., St George, K., & Walsh, T. J. (2013). Human rhinoviruses. *Clinical microbiology reviews*, *26*(1), 135–162. https://doi.org/10.1128/CMR.00077-12

Jandhyala, S. M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M., & Nageshwar Reddy, D. (2015). Role of the normal gut microbiota. *World journal of gastroenterology*, *21*(29), 8787–8803. https://doi.org/10.3748/wjg.v21.i29.8787

Janeway CA Jr, Travers P, Walport M, et al. Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science; 2001. Infectious agents and how they cause disease. Available from: https://www.ncbi.nlm.nih.gov/books/NBK27114/

Just, I., Selzer, J., von Eichel-Streiber, C., & Aktories, K. (1995). The low molecular mass GTP-binding protein Rho is affected by toxin A from C. difficile. *The Journal of clinical investigation*, *95*(3), 1026–1031. https://doi.org/10.1172/JCI117747

Kampf, G., Löffler, H., & Gastmeier, P. (2009). Hand hygiene for the prevention of nosocomial infections. *Deutsches Arzteblatt international*, *106*(40), 649–655. https://doi.org/10.3238/arztebl.2009.0649

Karjalainen, T., Barc, M. C., Collignon, A., Trollé, S., Boureau, H., Cotte-Laffitte, J., & Bourlioux, P. (1994). Cloning of a genetic determinant from C. difficile involved in adherence to tissue culture cells and mucus. *Infection and immunity*, *62*(10), 4347–4355.

Kelley, L., Mezulis, S., Yates, C. *et al.* The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10, 845–858 (2015). https://doi.org/10.1038/nprot.2015.053

Kilpatrick, A. M. & Altizer, S. (2010) Disease Ecology. *Nature Education Knowledge* 3(10):55

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R. M., & Centre for Mathematical Modelling of Infectious Diseases COVID-19 working group (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet. Infectious diseases*, *20*(5), 553–558. https://doi.org/10.1016/S1473-3099(20)30144-4

Lawley, T. D., Clare, S., Walker, A. W., Goulding, D., Stabler, R. A., Croucher, N., Mastroeni, P., Scott, P., Raisen, C., Mottram, L., Fairweather, N. F., Wren, B. W., Parkhill, J., & Dougan, G. (2009). Antibiotic treatment of C. difficile carrier mice triggers a supershedder state, spore-mediated transmission, and severe disease in immunocompromised hosts. *Infection and immunity*, *77*(9), 3661–3669. https://doi.org/10.1128/IAI.00558-09

Leitner, W. W., Wali, T., Kincaid, R., & Costero-Saint Denis, A. (2015). Arthropod Vectors and Disease Transmission: Translational Aspects. *PLoS neglected tropical diseases*, *9*(11), e0004107. https://doi.org/10.1371/journal.pntd.0004107

Lessa, F. C., Gould, C. V., & McDonald, L. C. (2012). Current status of C. difficile infection epidemiology. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, *55 Suppl 2*(Suppl 2), S65–S70. https://doi.org/10.1093/cid/cis319

Locher K. P. (2016). Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nature structural & molecular biology*, *23*(6), 487–493. https://doi.org/10.1038/nsmb.3216

Lozupone, C., Stombaugh, J., Gordon, J. *et al.* Diversity, stability and resilience of the human gut microbiota. *Nature* 489, 220–230 (2012). https://doi.org/10.1038/nature11550

Lucet, I., Feucht, A., Yudkin, M. D., & Errington, J. (2000). Direct interaction between the cell division protein FtsZ and the cell differentiation protein SpoIIE. *The EMBO journal*, *19*(7), 1467–1475. https://doi.org/10.1093/emboj/19.7.1467 Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual review of biophysics and biomolecular structure*, *29*, 291–325. https://doi.org/10.1146/annurev.biophys.29.1.291

McClung, RP, Roth, DM, Vigar, M, Roberts, VA, Kahler, AM, Cooley, LA, Hilborn, ED, Wade, TJ, Fullerton, KE, Yoder, JS & Hill, VR. Waterborne disease outbreaks associated with environmental and undetermined exposures to water — United States, 2013-2014. *Am J Transplant*. 2018; 18: 262- 267. https://doi.org/10.1111/ajt.14607

McDonald L. C. (2013). Editorial Commentary: looking to the future: vertical vs horizontal prevention of C. difficile infections. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, *57*(8), 1103–1105. https://doi.org/10.1093/cid/cit477

Mc Namara, K., Alzubaidi, H., & Jackson, J. K. (2019). Cardiovascular disease as a leading cause of death: how are pharmacists getting involved? *Integrated pharmacy research & practice*, *8*, 1–11. https://doi.org/10.2147/IPRP.S133088

Murphy, C. R., Avery, T. R., Dubberke, E. R., & Huang, S. S. (2012). Frequent hospital readmissions for C. difficile infection and the impact on estimates of hospital-associated C. difficile burden. *Infection control and hospital epidemiology*, *33*(1), 20–28. https://doi.org/10.1086/663209

Ottlinger, M. E., & Lin, S. (1988). C. difficile toxin B induces reorganization of actin, vinculin, and talin in cultured cells. *Experimental cell research*, *174*(1), 215–229. https://doi.org/10.1016/0014-4827(88)90156-5

Paredes-Sabja, D., Shen, A., & Sorg, J. A. (2014). C. difficile spore biology: sporulation, germination, and spore structural proteins. *Trends in microbiology*, *22*(7), 406–416. https://doi.org/10.1016/j.tim.2014.04.003

Pereira FC, Saujet L, Tomé AR, Serrano M, Monot M, Couture-Tosi E, et al. (2013) The Spore Differentiation Pathway in the Enteric Pathogen *C. difficile*. PLoS Genet 9(10): e1003782. https://doi.org/10.1371/journal.pgen.1003782

Phillips J. A. (2019). Chlamydia Infections. *Workplace health & safety*, *67*(7), 375–376. https://doi.org/10.1177/2165079919853590

Redelings, M. D., Sorvillo, F., & Mascola, L. (2007). Increase in C. difficile-related mortality rates, United States, 1999-2004. *Emerging infectious diseases*, *13*(9), 1417–1419. https://doi.org/10.3201/eid1309.061116

Reinke, C. M., & Messick, C. R. (1994). Update on C. difficile-induced colitis, Part 2. *American journal of hospital pharmacy*, *51*(15), 1892–1959.

Rinaldo, A., Bertuzzo, E., Blokesch, M., Mari, L., & Gatto, M. (2017). Modeling Key Drivers of Cholera Transmission Dynamics Provides New Perspectives for Parasitology. *Trends in parasitology*, *33*(8), 587–599. https://doi.org/10.1016/j.pt.2017.04.002

Rossati, A., Bargiacchi, O., Kroumova, V., Zaramella, M., Caputo, A., & Garavelli, P. L. (2016). Climate, environment and transmission of malaria. *Le infezioni in medicina*, *24*(2), 93–104.

Rupnik, M., Wilcox, M. & Gerding, D. *C. difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol* 7, 526–536 (2009). https://doi.org/10.1038/nrmicro2164

Saujet L, Pereira FC, Serrano M, Soutourina O, Monot M, Shelyakin PV, et al. (2013) Genome-Wide Analysis of Cell Type-Specific Gene Transcription during Spore Formation in *C. difficile*. PLoS Genet 9(10): e1003756. https://doi.org/10.1371/journal.pgen.1003756

Schorderet-Weber, S., Noack, S., Selzer, P. M., & Kaminsky, R. (2017). Blocking transmission of vector-borne diseases. *International journal for parasitology. Drugs and drug resistance*, *7*(1), 90–109. https://doi.org/10.1016/j.ijpddr.2017.01.004

Steglich, M., Hofmann, J. D., Helmecke, J., Sikorski, J., Spröer, C., Riedel, T., Bunk, B., Overmann, J., Neumann-Schaal, M., & Nübel, U. (2018). Convergent Loss of ABC Transporter Genes from *Clostridioides difficile* Genomes Is Associated with Impaired Tyrosine Uptake and *p*-Cresol Production.

Surawicz, C. M., & McFarland, L. V. (1999). Pseudomembranous colitis: causes and cures. *Digestion*, *60*(2), 91–100. https://doi.org/10.1159/000007633

Underwood, S., Guan, S., Vijayasubhash, V., Baines, S. D., Graham, L., Lewis, R. J., Wilcox, M. H., & Stephenson, K. (2009). Characterization of the sporulation initiation pathway of C. difficile and its role in toxin production. *Journal of bacteriology*, *191*(23), 7296–7305. https://doi.org/10.1128/JB.00882-09

Walters, B. A., Roberts, R., Stafford, R., & Seneviratne, E. (1983). Relapse of antibiotic associated colitis: endogenous persistence of C. difficile during vancomycin therapy. *Gut*, *24*(3), 206–212. https://doi.org/10.1136/gut.24.3.206

Yates, C. M., Filippis, I., Kelley, L. A., & Sternberg, M. J. (2014). SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. *Journal of molecular biology*, *426*(14), 2692–2701. https://doi.org/10.1016/j.jmb.2014.04.026

Zeng, W., Vorst, K., Brown, W., Marks, B. P., Jeong, S., Pérez-Rodríguez, F., & Ryser, E. T. (2014). Growth of Escherichia coli O157:H7 and Listeria monocytogenes in packaged fresh-cut romaine mix at fluctuating temperatures during commercial transport, retail storage, and display. *Journal of food protection*, *77*(2), 197–206. https://doi.org/10.4315/0362-028X.JFP-13-117

Zhang, G., Chen, Y., Hu, L., Melka, D., Wang, H., Laasri, A., Brown, E. W., Strain, E., Allard, M., Bunning, V. K., Parish, M., Musser, S. M., & Hammack, T. S. (2018). Survey of Foodborne Pathogens, Aerobic Plate Counts, Total Coliform Counts, and Escherichia coli Counts in Leafy Greens, Sprouts, and Melons Marketed in the United States. *Journal of food protection*, *81*(3), 400–411. https://doi.org/10.4315/0362-028X.JFP-17-253

VI. APPENDIX

STVLILLTSIVGGATMGASSGVIIGVASILNNMTSAVYMGIYSFSGLISGAFNKINKYFCIL
GYILSWTIIYLYTSGITSNMMQLRDILLGCLIVLVLPERLFNKIEKLIKSNVASNEIVYDYI
MRSKNLTNSRLNSIYKTYDDLADTFDKIREKDKVLDQRDIANVIDMIHNDECKSCSMRR
MCWESRFNHTYTMVYNILEKIEEKGELSLNDIPKNFRKECMKPESIVKISNHYYKMFVL
DYDWSVKFSESRKLIANQIRSISKSIKSLSQDLEGDIMLDIEKEKNIYEQLERYDITVDKVS
YLTKSNSEFEISIEKKTCHDGCMCEDKIVNIISDLVGENMSVRKIGCHCLGGKCKATFVK
SQKYKAVTEVSAMSRDGHILCGDNYTYMEINDGKYMMAISDGMGKGKKAYEESSATI
DILEKMIDAKIKDEIVIDTINNMLLLKSSEEMFSTLDLGILDLKRGCLETIKMGACSTYIKR
EDGEVDLISSSSLPVGILSDVKIDRKNVKVKEGDYVIMVSDGIVDAGRNNNLGDNWLIY
FLKNIETTNPKEISNLILDRALELQALQI

Figure A 1. FASTA Sequence *C. difficile* SpoIIE, UA001

MEKAERRVNGPMAGQALEKLQSFFNRGTKLVTHHLHSLFFYKGFIYVVIGFLLGRAFIL
SEVLPFALPFFGAMLLIRRDKAFYAVLAVLAGALTISPKHSLLILAALLAFFVFSKVAAFI
TDDRVKALPIVVFFSMAAARAGFVYAQNGVFTTYDYVMAIVEAGLSFILTLIFLQSLPIF
TVKKVKQSLKIEEIICFMILIASVLTGLAGLSYQGMQAEHILARYVVLSFSFIGGASIGC
TVGVVTGLILGLANIGNLYQMSLLAFSGLLGGLLKEGKKAGAAIGLIVGSLLISLYGEGS
AGLMTTLYESLIAVCLFLLTPQSITRKVARYIPGTVEHLQEQQQYARKIRDVTAQKVDQF
SNVFHALSESFATFYQASDEQTDDSEVDLFLSKITEHSCQTCYKKNRCWVQNFDKTYDL
MKQVMLETEEKEYASNRRLKKEFQQYCSKSKQVEELIEDELAHHHAHLTLKKKVQDSR
RLVAEQLLGVSEVMADFSREIKREREQHFLQEEQIIEALQHFGIEIQHVEIYSLEQGNIDIE
MTIPFSGHGESEKIIAPMLSDILEEQILVKAEQHSPHPNGYSHVAFGSTKSYRVSTGAAHA
AKGGGLVSGDSYSMMELGARKYAAAISDGMGNGARAHFESNETIKLLEKILESGIDEKI
AIKTINSILSLRTTDEIYSTLDLSIIDLQDASCKFLKVGSTPSFIKRGDQVMKVQASNLPI
GIINEFDVEVVSEQLKAGDLLIMMSDGIFEGPKHVENHDLWMKRKMKGLKTNDPQEIA
DLLMEEVIRTRSGQIEDDMTVVVVRIDHNTPKWASIPVPAIFQNKQEIS

Figure A 2. FASTA Sequence *B. subtilis* SpoIIE

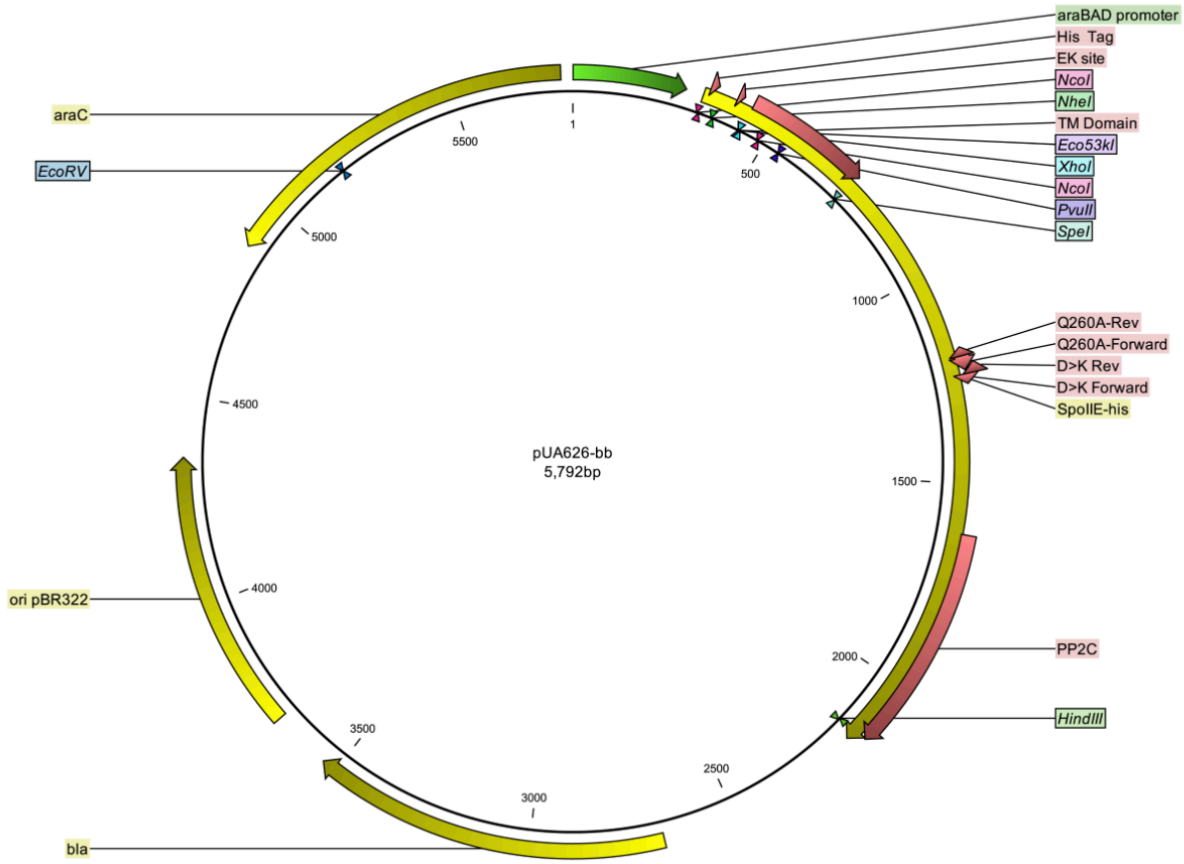

Figure A 3. Plasmid map pUA626 with SpoIIE: close-up

TM region	Linker region	Alpha helix	Regulatory | Catalytic

Figure A 4. Depiction of SpoIIE domains