East Tennessee State University
Digital Commons @ East Tennessee State University

Electronic Theses and Dissertations | Student Works

5-2014

Characterization of TonB in Rhizobium leguminosarum ATCC 14479

Brian D. Hill
East Tennessee State University

Follow this and additional works at: https://dc.etsu.edu/etd

 Part of the Microbial Physiology Commons

Recommended Citation

Hill, Brian D., "Characterization of TonB in Rhizobium leguminosarum ATCC 14479" (2014). *Electronic Theses and Dissertations.* Paper 2379. https://dc.etsu.edu/etd/2379

This Thesis - unrestricted is brought to you for free and open access by the Student Works at Digital Commons @ East Tennessee State University. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ East Tennessee State University. For more information, please contact digilib@etsu.edu.

Characterization of TonB in *Rhizobium leguminosarum* ATCC 14479

A thesis

presented to

the faculty of the Department of Health Sciences

East Tennessee State University

In partial fulfillment

of the requirement for the degree

Master of Science in Biology

---

by

Brian D. Hill

May 2014

Dr. Ranjan Chakraborty, Chair

Dr. Christopher Pritchett

Dr. Dhirendra Kumar

Keywords: *Rhizobium leguminosarum*, TonB, vicibactin

ABSTRACT

Characterization of TonB in *Rhizobium leguminosarum* ATCC 14479

by

Brian Hill

Rhizobium leguminosarum is a gram-negative soil bacterium that requires iron for survival. However, iron becomes insoluble in the presence of oxygen at physiological pH. In response, *Rhizobia* species have used siderophore mediated iron transport systems to meet their iron requirements. *R. leguminosarum* ATCC 14479 produces the trihydroxymate siderophore vicibactin and we hypothesize that the import of the ferric iron-vicibactin complex is energized by the TonB-ExbB-ExbD system. Here, we have identified a putative *tonB* gene. A *tonB* mutant was created and compared with wild type in its ability to transport ^{55}Fe-vicibactin. Also, the putative TonB of *R. leguminosarum* ATCC 14479 is interesting due to its estimated size compared to the TonB of *E. coli.* Many groups have attempted structural analysis of the C-terminus of TonB in *E. coli* with inconsistent results. We were successful in expressing 2 different sized TonB C-terminals (120 and 200 amino acids) using pET17b in *E. coli.*

ACKNOWLEDGEMENTS

I would first like to thank my committee chair, Dr. Ranjan Chakraborty, for the opportunity to receive his mentoring. He has been unwavering in his support of my work and always been gracious with his time. He has been much more than a mentor; he has been a friend. I would like to thank the other members of my committee, Dr. Dhirendra Kumar and Dr. Christopher Pritchett, for their time, support, and contribution to my work. I would also like to thank the faculty and employees of the Health Sciences Department and my fellow graduate students, past and present. Finally, I would like to thank my family and friends for their support during my pursuit of my degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Iron Abundance and Role

Analyses of the Earth's upper crust have determined that iron is the fourth most common element, ranking behind oxygen, silicon, and aluminum (Rudnick and Gao 2003). Due to iron's relative abundance, microorganisms, along with almost all other life forms, have exploited the redox potential that exist between Fe(II)/Fe(III) thus harvesting energy by using iron as an electron carrier. For example, in aerobic and anaerobic conditions, Fe-oxidizing microorganisms generate energy by oxidizing ferrous iron (Weber et al. 2006). Meanwhile, Fe-reducing microorganisms use ferric iron as a terminal electron acceptor in anaerobic conditions (Weber et al. 2006). Other microorganisms incorporate iron into the prosthetic moieties of proteins that contribute to metabolic pathways such as photosynthesis and respiration, nitrogen fixation, methanogenesis, the regulation of genes, and the biosynthesis of DNA (Andrews et al. 2003). It has also been proposed that Fe(II) acted similarly to Mg(II) ions' current role in the folding and functioning of large RNAs in early earth, prior to the great oxidation event catalyzed by photosynthesizing cyanobacteria (Athavale et al. 2012).

Iron Availability

Because iron participates in many major biological pathways, its acquisition is necessary and inherently easier when the metal is found in a soluble form. Although at pH of 4 or below iron remains in a soluble ferrous state even when confronted by oxic conditions, many organisms inhabit environments of a more neutral pH (Weber et al.

2006). In these oxic physiological conditions (approximate pH of 7), insoluble ferric iron predominates while the concentration of environmental soluble ferrous iron has been estimated at 10^{-18} M (Andrews et al. 2003). Similarly, pathogenic microorganisms are confronted with a free serum iron concentration of 10^{-24} M in hosts that produce lactoferrin and transferrin (Miethke and Marahiel 2007). Therefore, many microorganisms have had to transition from an environment well supplied with readily available ferrous iron that could be easily acquired and quickly incorporated in its reduced state into prosthetic groups to an environment where iron acquisition systems are necessary to transport insoluble ferric iron intracellularly. Once internalized, the reduction of ferric iron is required for biologically functional incorporation.

Iron Scavenging Systems

Because iron is found environmentally and biologically in both free and complexed forms, microorganisms have evolved scavenging mechanisms that function to acquire iron from the available iron sources. Direct iron scavenging systems require a specialized outer membrane receptor that recognizes specific iron sources such as transferrin, lactoferrin, and heme (Miethke and Marahiel 2007). An example of a direct uptake system is found in *Pseudomonas aeruginosa* and involves the Fur regulated *phuR-phuSTUVW* system, where PhuR is an outer membrane receptor specific for heme and the *phuSTUVW* operon encodes an ABC type transporter (Ochsner et al. 2000). Another direct uptake system is the bipartite heme receptors that have been identified in Neisseria species, which consist of a TonB-dependent outer membrane receptor that is specific for heme (Anzaldi and Skaar 2010). These particular strategies limit an organism to

particular environments where these iron sources are available, and restrict the organisms from colonizing areas where iron sources are available but not incorporated into these specific proteins. The 2 major indirect acquisition methods are the hemophore-mediated heme uptake system and the siderophore-mediated iron uptake system. The HasA hemophore of *Pseudomonas aeruginosa* is a heme binding protein that is secreted into the extracellular environment and once bound to heme, recognized by an outer membrane receptor (HasR) and transported into the periplasm by either TonB or the TonB-like protein HasB (Anzaldi and Skaar 2010). Finally, the most prevalent, effective, and least ecologically restrictive system that microorganisms employ is the siderophore-mediated iron acquisition system.

Siderophores

Siderophores are low molecular weight molecules (< 1 kD) with a high affinity for ferric iron (Neilands 1981). Siderophores are produced and secreted by bacteria, fungi (e.g. ferrichrome secreted by *Ustilage sphaerogena)*, and a small number of plant species in order to make available iron that has been rendered unavailable by the formation of insoluble ferric complexes (Miethke and Marahiel 2007). For example, one known class of plant siderophores called mugineic acid-family phytosiderophores are secreted from the roots when confronted with iron limiting conditions (Shojima et al. 1990). Bacteria, however, secrete three main classes of siderophores: catecholates (catecholates and phenolates), hydroxymates, and carboxylates (Miethke and Marahiel 2007). However, it has been found that siderophores can contain characteristics of more than one class and can therefore be classified as “mixed”. The class determines the

physiochemical properties of the molecule. Siderophores are often characterized by their stability constant, which is used to evaluate the siderophore's affinity for ferric iron and its pK_a value, both of which determine the siderophore's ability to chelate Fe(III) by evaluating the siderophore donor groups' ability to protonate within a specific range of pH (Miethke and Marahiel 2007). In general, catecholate-type siderophores exhibit the highest stability constant and thus possess the highest affinity for ferric iron (Miethke and Marahiel 2007). For example, enterobactin, the catecholate-type siderophoreproduced by *E. coli,* has a stability constant of 10^{52} (Loomis and Raymond, 1991). However, the pK_a values must also be taken into consideration when making these determinations. For instance, catecholate-type, hydroxamate-type, and carboxylate-type siderophores have pK_a values of 6.5-8, 8-9, and 3.5-5, respectively (Meithke and Marahiel 2007). Thus, carboxylate-type siderophores may provide a competitive advantage against catecholate-type siderophores when secreted into extremely acidic conditions even though they inherently possess a lower stability constant and lesser affinity.

Siderophore Specific Outer Membrane Transporters

Once siderophores are biosynthesized, secreted into the surrounding environment, and successfully bound to ferric iron, they must be transported across the cell wall and periplasmic membrane to the cytoplasm. Because iron-siderophore complexes are too large to be passively transported across the outer membrane (OM) of bacteria and are often species specific, a transporter that binds the unique iron-siderophore complex with high affinity must be present in the outer membrane. Although these OM transports are specific for a particular iron-siderophore complex, they all possess a similar structural

conformation consisting of a 22 β-stranded barrel across the OM to the periplasmic space and a N-terminal globular domain that acts to "plug", blocking an otherwise unobstructed pore inside of the cylinder (Chimento et al. 2005). These conformations have been confirmed by the crystal structures of FhuA (ferric hydroxamate uptake transporter), FepA (ferric-enterobactin permease), and FecA (ferric-citrate protein), all of which contained the barrel and plug conformation (Bertini et al. 2007). Tight binding of the iron-siderophore is achieved by the contribution of residues from the globular plug domain along with residues that are exposed in the lumen of the barrel (Chimento et al. 2005). Once the iron-sideophore complex has bound its specific outer membrane transporter, a conformational change resulting in the disordering of a conserved motif located within the N-terminus of the outer membrane transporter occurs (Gumbart et al. 2007). This conserved motif is called the TonB box. Because the binding of the "TonB box" by TonB followed by the transduction of energy by TonB is required by these outer membrane transporters to actively transport the iron-siderophore complex intracellularly, they are often referred to as TonB-dependent transporters (TBDT).

TonB-ExbB-ExbD Complex

Structurally, Gram-negative bacteria are composed of an outer membrane, periplasmic space, and an inner membrane or cytoplasmic membrane. The outer membrane is devoid of the energy generating proton motive force (pmf) that is present in the inner membrane. The pmf is critical in providing an electrochemical gradient capable of generating energy to perform cellular work. Because the size of the iron-siderophore complex is too large to diffuse through porins and the concentration gradient is not

receptive to iron acquisition (10^{-18} M extracelluar concentration and $10^{-6} – 10^{-7}$ M intracellular concentration (Andrews et al. 2003)), the iron-siderophore complex must be actively transported into the cell. This poses a dilemma for Gram-negative organisms since the TBDT does not have an available energy source in the outer membrane or the periplasmic space to energize transport. The answer lies in the TonB-ExbB-ExbD complex present in the cytoplasmic membrane. The TonB-ExbB-ExbD complex has been most extensively researched in *E. coli*. Because ExbB and ExbD are homologous to the MotA and MotB proteins that couple the pmf to produce movement in bacterial flagellum, it is believed that TonB-ExbB-ExbD complex couples the energy generated by the pmf (Zhai et al. 2003). The direct contact of TonB to the TonB box of the TBDT somehow transduces the energy necessary to facilitate the transport of the iron-siderophore complex (James et al. 2008).

ExbD is the smallest of the 3 proteins that make up the inner membrane complex. ExbD contains 1 transmembrane domain (TMD), a short cytoplasmic tail, and a long C-terminus, which allows for extension into the periplasmic space (Ollis et al. 2012). It has recently been proposed that ExbD assist the energizing process by facilitating three stages of conformational changes within TonB (Ollis et al. 2012). The first stage involves no interaction between ExbD and TonB; the interaction of residues 62-141of ExbD and TonB characterize the second stage, and finally, the last stage involves the pmf driven interaction between residues 42-61 of ExbD and TonB (Ollis et al. 2012).

ExbB contains 3 TMDs with the N-terminus residing in the cytoplasm and the C-terminus residing in the periplasm (Baker and Postle 2013). It has recently been revealed that the first, second, and third TMDs are made up of residues 22-42, 132-152, and 178-

198, respectively (Baker and Postle 2013). In the same study, amino acid substitutions to residues proposed for signal and energy transduction showed that the TMDs of ExbB do not participate in the proton pathway (Baker and Postle 2013). Although ExbB is proposed to not be a part of the proton translocation, it has long been proposed to be the scaffold for the entire TonB complex. Previous studies have shown through radiolabelling that the overall cellular composition of ExbB:ExbD:TonB is present in a ratio of 7:2:1 (Higgs et al. 2002). However, recent studies have shown that for every 1 ExbD protein required to produce the energizing complex that 6 ExbB proteins are present to form the foundation (Pramanik et al. 2011). The ratio of TonB was not analyzed in the aforementioned study.

The final component of the complex is TonB. Members of the TonB family of proteins have been identified within many different Gram-negative species and all TonB proteins share a similar topology: a single TMD, a short cytoplasmic N-terminal domain, a proline-rich region, and a periplasmic C-terminal domain (Postle and Larsen 2007). The N-terminal domain of TonB (residues 1-33) contains a sequence called the "SHLS" motif that is conserved among the TonB members (Larsen and Postle 2001). A recent study concluded that the histidine residue within this motif was the only residue within this motif that was critical for a functional TonB (Swayne and Postle 2011). This study also showed that an asparagine was the only amino acid capable of replacing histidine to produce a functional TonB. This is important because asparagine is a nonprotonatable amino acid, so histidine serves only a structural function within this motif and not a possible amino acid within the proton translocation pathway (Swayne and Postle 2011). The proline-rich region of TonB (34-154) contains 1 proline for every 3 amino acids. It

is thought to provide the rigidity necessary to extend across the periplasmic space in order to contact the TonB box of the TBDT. A study by Larsen showed that a reduction in the size of the proline rich region (Δ66-100) did not affect the ability of TonB to transport phi 80 and provided evidence to support the idea that the proline rich region provides solely structural support and does not have a role in energy transduction (Larsen et al. 1993). However, in the same study a reduction in transport was observed when cells were grown in conditions that resulted in a physiological expansion of the periplasmic space (Larsen et al. 1993). The C-terminal domain (CTD) of TonB (155-239) resides in the periplasm and is essential for binding and the energy transduction process. Solution structures of the CTD (103-239) have produced ordered configurations that consist of 2 α helices and 4 β strands (Peacock et al. 2005). During TonB binding to the TonB box of the TBDT BtuB, the cocrystal structures produced showed that the TonB box of BtuB relocated the fourth β strand of the CTD and inserted itself forming a stabilized β sheet that provides the pathway for energy transfer (Shultis et al. 2006). A study investigating the role of Q160 in transport revealed that single, double, and triple deletions encompassing Q160 reduced transport by 30%, while quadruple deletions and sextuple deletion resulted in a 70% and 100% reduction in activity, respectively (Vakharia-Rao et al. 2007). Many groups have attempted to produce structures of TonB and these results have often been conflicting. A CTD consisting of residues 164-239 (Chang et al. 2001) and a CTD consisting of residues 148-239 (Koedding et al. 2005) both produced dimers. Also, studies using a TonB-ToxR fusion protein produced evidence that indeed TonB acted as a dimer *in vivo* (Ghosh and Postle 2005). However, Peacock and colleagues examined residues 103-239 using NMR and reported that the CTD formed a monomer,

which was quickly backed up by a study reporting that TonB was capable of interacting with TBDTs as a monomer (Koedding et al. 2005). Much work is still need to be done in order to determine the structure and organization of TonB.

Rhizobium leguminosarum

Rhizobium leguminosarum is a Gram-negative, motile, aerobic bacillus that is a member of the class alpha-proteobacteria. *R. leguminosarum* can be found in the soil in a free living form or more importantly, in the roots of leguminous plants where they convert atmospheric nitrogen gas into ammonia via the process called nitrogen fixation. The process begins with the secretion of a mixture of flavonoid inducers by the host plant in response to low nitrogen soil conditions, which in turn activate the transcriptional activator NodD to trigger the transcription of the *nod* operons (Long 1996). NodD effectively binds flavonoids and induces the transcription of the *nodABCIJ*, *nodFEL*, *nodMNT*, and *nodO* operons by binding of the *nod* box (Yang et al. 2012). The general structure of the Nod factors is produced by the actions of NodA, NodB, and NodC, while the modifications to produce a species-specific factor lie in the actions of the remaining *nod* genes (Walker and Downie 2000). The infection process also entails quorum sensing of acyl homoserine lactones (AHLs) that may assist in growth, survival during stationary phase, and host-specific associations (McAnulla et al. 2007). Quorum sensing along with the nodulation factors are critical in the formation of an infection thread and successful nodule formation (Walker and Downie 2000).

Iron Regulation in *Rhizobium leguminosarum*

In the presence of hydrogen peroxide, ferrous and ferric iron undergoes the Fenton reaction and produces highly reactive oxygen species in the form of hydroxyl radicals and ferryl iron (Brennelsen et al. 1998). Sugars, phospholipids, organic acids, amino acids, and DNA bases can be targeted and damaged by these powerful oxidizing radicals (Prousek 2007). Due to its necessity at low concentration and its toxicity as intracellular concentrations begin to elevate, iron homeostasis must be tightly regulated within microorganisms. One of the most studied regulators of iron is the "ferric uptake regulation" gene or *fur* (Guerinot 1994). In Gram-negative bacteria, Fur is the most widely employed iron regulator protein. Fur binds to the conserved sequences that overlap promoter regions called "Fur boxes", and represses the transcription of over 90 genes that are involved in siderophore biosynthesis and uptake (Hantke 2001). Although Fur has also been found to regulate iron concentrations within some Gram-positive bacteria and cyanobacteria, Gram-positive bacteria often use Dtx to regulate iron concentrations (Andrews et al., 2003). Along with siderophore biosynthesis and uptake, Fur has been shown to regulate genes involved in acid shock response, detoxification of oxygen radicals, the production of toxins and other virulence factors, quorum sensing, and the regulation of metabolic pathways (Agarwal et al. 2008).

In *Rhizobium leguminosarum*, Fur is replaced by a different global regulator of iron that shows no homology to Fur called RirA, (Todd et al. 2002). A *rirA*- mutant was shown to transcriptionally regulate *rpoI*, the ECF factor that is involved in regulating the vicibactin synthesis operons *vbsGSO* and *vbsADL*, *fhuA* and *fhuCDB* (a TBDT and an ATPase, periplasmic binding protein, and inner membrane transporter, respectively),

orf1-tonB operon that encodes a putative TonB family protein, and *hmuPSTUV* that encodes an ABC-type transporter required for haem uptake (Todd et al. 2002). RirA is thought to bind a conserved region called an iron responsive operator (IRO) or more specifically in this case, the RirA box (Yeoman et al. 2004). There exist another iron regulator in *R. leguminosarum* called IrrA, which binds iron control elements (ICE) motifs and is regulated in iron deficient conditions by RirA but has been shown to regulate *rirA* expression in iron replete conditions (Todd et al. 2006). IrrA is shown to induce the transcription of genes involved in the synthesis of FeS clusters during iron replete conditions (Todd et al. 2006). Also, iron regulation in *R. leguminosarum* seems to depend on the interplay between these 2 regulators as operons may be regulated exclusively by RirA, IrrA, or by a combination of the 2 when both IRO and ICE motifs are present (Todd et al. 2006).

Vicibactin

Vicibactin is a cyclic trihydroxymate siderophore produced by *R. leguminosarum* (Carson et al. 1994). As mentioned earlier, the transcription of operons *vbsGSO* and *vbsADL* are controlled by RpoI (Todd et al. 2002). However, the other 2 genes involved in vicibactin synthesis, *vbsC* and *vbsP*, are not under RpoI control (Carter et al. 2002). Vicibactin is synthesized as a monomer through a multistep process involving modifications made by VbsO, VbsP, VbsA, VbsL, VbsC, VbsS and culminating in VbsS catalyzing an esterification reaction that results in trimerization and cyclization of the modified monomer (Carter et al. 2002). While proposals of how vicibactin is synthesized

within the cell exist, how vicibactin is exported extracellularly and how and where vicibactin undergoes degradation remain unclear.

Identification of a Putative *tonB* in *Rhizobium leguminosarum*

In 2002 Johnston and colleagues identified a putative *tonB* gene while performing transposon-mediated mutagenesis and screening phenotypically for vicibactin overexpressers on CAS media (Wexler et al. 2002). The gene encoded a putative protein estimated to be 440 amino acids long (Wexler et al. 2002) that shared similarity with the much smaller TonB of *E. coli*, which is known to be 239 amino acids long (Postle and Good 1983). Subsequent analysis showed that this putative TonB protein contained an "SHLS" motif that is necessary for binding ExbB in the TonB-ExbB-ExbD inner membrane complex and a high percentage (11%) of proline with several Pro-Glu and Pro-Lys repeats located between the putative N-terminal and C-terminal (Wexler et al. 2002).

It was also found that this putative *tonb* was cotranscribed with the upstream *orf1* and the promoter region was located approximately 80 bp upstream of the *orf1*'s putative translational start site (Wexler et al. 2002). This fragment, *orf1-tonb*, was then used to fully complement the vicibactin overexpresser (Wexler et al. 2002).

Rhizobium leguminosarum ATCC 14479

Rhizobium leguminosarum ATCC 14479 was isolated at Arlington Farms, VA and purchased from the American Type Culture Collection. It was determined to be most effective at nodulating the roots of red clover, *Trifolium praetense*. *R. leguminosarum*

ATCC 14479 may also be known as USDA 2046, DSM 6040, and *Rhizobium trifolii* Dangeard 1926 (Ramirez-Bahena et al. 2008).

In our lab, *R. leguminosarum* ATCC 14479 was found to produce vicibactin. (Wright et al. 2013).

Present Work

We have shown that *Rhizobium leguminosarum* ATCC 14479 produces the trihydroxymate siderophore vicibactin when exposed to iron-deficient conditions (Wright et al 2013). We hypothesize that the import of the ferric iron-vicibactin complex is energized by the TonB-ExbB-ExbD system that is found within other gram-negative bacteria. In this study we have identified a putative *tonB* gene. A *tonB* knockout mutant was created and compared with wild type ATCC 14479 in its ability to grow in complex media, high iron, and low iron minimal media. We also compared the knockout mutant with wild type for the ability to transport Fe^{55} loaded vicibactin. We were able to restore wild type growth and Fe^{55}-vicibactin in the *tonB* mutant by mobilizing a plasmid containing the promoter-*orf1-tonB* sequence. Transport was partially restored in the *tonB* mutant by using 2 plasmids lacking the upstream *orf1* and using 2 different ribosomal binding sites and starts sites for *tonB*. This potentially alters the predicted length of the putative TonB to 353 amino acids. Also, the TonB of *R. leguminosarum* is interesting due to its predicted size, 450 amino acids, compared to the TonB of *E. coli*, which is 239 amino acids. Many groups have attempted structural analysis of full length TonB and the C-terminus of TonB in *E. coli*. The results of these analyses have been inconsistent, producing a variety of crystallographic and NMR structures. We attempt to express full

length TonB and 2 different sized TonB C-terminal domains (120 and 200 amino acids long) using pET17b in the heterologous species *E. coli.*

CHAPTER 3

MATERIALS AND METHODS

Bacterial Strains

The rhizobia strain used in this study was *Rhizobium leguminosarum* ATCC 14479. The *E. coli* strains used in this study were DH5α, SM10λ*pir*, and BL21 (DE3) (Novagen). Strains used in this work can be seen in Table 1.

Table 1. Strains and Plasmids

Strains	Characteristics	References
E. coli		
DH5α	used for general cloning and triparental conjugation	
SM10λpir	used during biparental conjugation to mobilize pEXTB into ATCC 14479	
BL21(DE3)	used to express pET120 and pET200 products	
R. leguminosarum ATCC 14479		
	wildtype strain	
RLΔTB	ATCC 14479 with an in-frame deletion of *tonB*	This work
Plasmids		
pUC19	general cloning vector; Amp^R	Messing et al. 1983
pUCTB	pUC19 containing *R. leguminosarum tonB* PCR product	This work
pUCTBS	pUC19 containing *R. leguminosarum* SOE product with in-frame deletion of *tonB*.	This work
pUCTBFR	pUC19 containing wildtype F1-R1 sequence	This work
pEX18Gm	suicide vector; Gm^R	Schweizer et al. 1998
pEXTBS	pEX18Gm containing *R. leguminosarum* SOE product with in-frame deletion of *tonB*.	This work
pMR10	RK2 based low copy plasmid used for complementation and expression; Km^r	Mohr and Roberts, unpublished
pMRTBC	pMR10 containing F1R1 fragment containing *orf1-tonB* used for complementation	This work
pMRTP1	pMR10 containing F1R1 fragment without *orf1* and 1st potential RBS	This work
pMRTP2	pMR10 containing F1R1 fragment without *orf1* and 2nd potential RBS	This work
pET17b	pUC based expression vector; Amp^R	Novagen
pET120	pET17b containing 120 amino acid long C-terminus	This work
pET200	pET17b containing 200 amino acid long C-terminus	This work
pET457	pET17b containing full-length *R. leguminosarum* TonB	This work
pBIO1187	pLAFR1-based cosmid, contains vbs genes plus *rpoI*; Tet^R	Yeoman et al. 2002
pRK2013	helper plasmid for triparental conjugation; Km^R	Helinski et al. 1980

Growth Conditions

Three types of media were used to culture *Rhizobium leguminosarum* ATCC 14479: Congo Red agar (CR), Yeast Mannitol Broth (YMB), and Modified Manhart and Wong broth (MMW).

The Congo Red media is a modified yeast mannitol agar with the addition of Congo Red dye (Jadhav and Desai 1996) and contains the following ingredients (w/v): 1% mannitol, 0.05% K_2HPO_4, 0.02% $MgSO_4*7H_20$, 0.01% NaCl, 0.1% yeast extract, 2.5 x 10^{-5} % Congo Red dye, and 3% Bacto-agar. Prior to autoclaving, the aforementioned percent volume of Congo Red dye was added to the media and the media was adjusted to the pH of 6.8 using 6M NaOH. Congo Red dye is slowly absorbed by *Rhizobium* species allowing the colonies to maintain a white or slightly pink coloration, but quickly taken up by contaminating organisms (Kneen and Larue 1983). This provided a mechanism of differentiation capable of distinguishing contaminating organisms.

Yeast Mannitol broth was used to routinely culture rhizobia strains when iron concentration was not of importance. YMB contains the following ingredients (w/v): 1% mannitol, 0.05% K_2HPO_4, 0.02% $MgSO_4*7H_20$, 0.01% NaCl, and 0.1% yeast extract. Prior to autoclaving, the media was adjusted to the pH of 6.8 using 6M NaOH.

Modified Manhart and Wong media (Manhart and Wong 1979) was used when there was an attempt to control iron concentration. MMW media consist of a basal media and a vitamin solution. Because many of the metals and vitamins present in the vitamin solution are present in trace amounts, the vitamin solution was concentrated 1000 times. The vitamin solution was then filter sterilized, wrapped in aluminum foil to prevent the degradation of light sensitive ingredients, and kept at 4□C. The vitamin solution was

added to the basal media prior to inoculation. The composition of MMW can be found in Appendix A . Prior to autoclaving, the basal media was adjusted to the pH of 6.8 using 6M NaOH.

CR agar plates inoculated with *R. leguminosarum* ATCC 14479 were grown at 30□°C. YMB and MMW broths inoculated with *R. leguminosarum* ATCC 14479 were grown at 30□°C on a shaker deck set at 200 rpm.

Antibiotic concentrations used for *R. leguminosarum* ATCC 14479 were as follows: nalidixic acid 10□μg/□l (Subscr. Na10); penicillin G 50 μg/□l (Subscr. Pg50); tetracycline 10 μg/ml (Subscr. Tc10); gentamicin 15 μg/ml (Subscr. Gm10); and kanamycin 50 μg/ml (Subscr. Km50).

All *E. coli* strains were grown in modified Luria-Bertani broth (LB). LB broth contains the following ingredients (g%): 1% Tryptone, 0.5% Yeast Extract, and 0.5% NaCl. When making LB agar plates, 1.5% agar was added. Prior to autoclaving, the media was adjusted to the pH of 7.5 using 6M NaOH.

Antibiotic concentrations used for *E. coli* were as follows: tetracycline 10 μg/ml (Subscr. Tc10); gentamicin 15 μg/ml (Subscr. Gm15); ampicillin 100 μg/ml (Subscr. Ap100); kanamycin 50μg/ml (Subscr. Km50).

Primers

The primers used during this work are found in Table 2.

Table 2. Primers

Primers	
Name	**Sequence**
TonBF	5’ GATGAACCCCGGCGGGCATG 3’
TonBR	5’ CTGCGGTGGGCTGTACAGC 3’
SOEF1	5’ GCGCGGTACCCAGCTGCGCGGCGATATCG 3’
SOEF2	5’ CGAGGGCTAAGAATACGCCGGAAGTTCAGGTCACTAGATAATATG 3’
SOER1	5’ GCGCTCTAGAGACCGTGTGCGAGGAATAGAG 3’
SOER2	5’ CATATTATCTAGTGACCTGAACTTCCGGCGTATTCTTAGCCCTCG 3’
TonBFL	5’ GCGCAAGCTTGATGCATCCCGGCCACGAGC 3’
TonBPrCt	5’ GCGCAAGCTTGATGCCTGATATCGCGCAGGC 3’
TonBCT	5’ GCGCAAGCTTGATGGTGAACGGCCAGGACG 3’
TonBR17B	5’ GCGCGAATTCGATGCCTGATATCGCGCAGG 3’
TBProF1	5’ GATATGAAGCCGCCTGGCGGGAGGGCTAAGAATACGCCGG 3’
TBProR1	5’ CCGGCGTATTCTTAGCCCTCCCGCCAGGCGGCTTCATATC 3’
TBProF2	5’ GATATGAAGCCGCCTGGCGGGCAGGAAGACGAGAGGAAGC 3’
TBProR2	5’ GCTTCCTCTCGTCTTCCTGCCCGCCAGGCGGCTTCATATC 3’

Genomic Extraction

Genomic extractions were performed according to Chopra et al. (1998). *R. leguminosarum* ATCC 14479 was grown in a 3 mL $MMW_{Na10\ Pg50}$ at 30□°C on a shaker deck set to 200 rpm. Due to ATCC 14479’s slow growth, naldixic acid was used to prevent any possible *E. coli* contamination while penicillin G was used to prevent contamination by Gram-positive species. Chopra’s protocol was modified to include 3 additional 0.85% NaCl washes between the initial pelleting to remove growth media and resuspension in TNE. These sterile saline washes were performed to remove excess exopolysaccharides from the cell pellet that may interfere with steps performed later in the protocol. All samples were ran on 1% TBE to verify extraction of genomic DNA.

Detection of *tonB* via PCR

The primers TonBF and TonBR (Table 2) were designed by the Johnston group for the initial identification of a putative *tonb* in *R. leguminosarum* WSM 2304 (Wexler et al. 2002). In this study, these primers were modified by the addition of a KpnI and XbaI restriction site to the 5' end of TonBF and TonBR, respectively. This allowed for quick digestion and ligation into pUC19 for sequencing.

Splicing by Overlap Extension

Splicing by overlap extension (SOE) is a process by which 2 PCR reactions are subsequently used to engineer a product that contains the 1000 base pair "upstream" region of a particular gene of interest and the 1000 base pair "downstream" region fused together, thus eliminating the targeted gene (Horton et al. 1990). Two sets of primers were developed to carry out this protocol for *tonB* in ATCC 14479 (Fig. 1). The first set of primers SOEF1-SOER2 was used to amplify the 1000 base pair (bp) upstream flanking region of *tonB*. The second set of primers SOEF2-SOER1 was used to amplify the 1000 bp downstream flanking region. Because the genome of ATCC 14479 has not been sequenced, these primers were designed by using the WSM 2304 genome, which is available through the KEGG database. Once the 2 products were produced via PCR, a subsequent PCR reaction was used to generate a 2000 base pair product with an in-frame deletion of *tonB*. In this reaction, equal amounts of each product were used as template and SOEF1 and SOER1 were used as primers. The reaction generated the desired SOE construct, which was subsequently ligated into pUC19 and sent for sequencing.

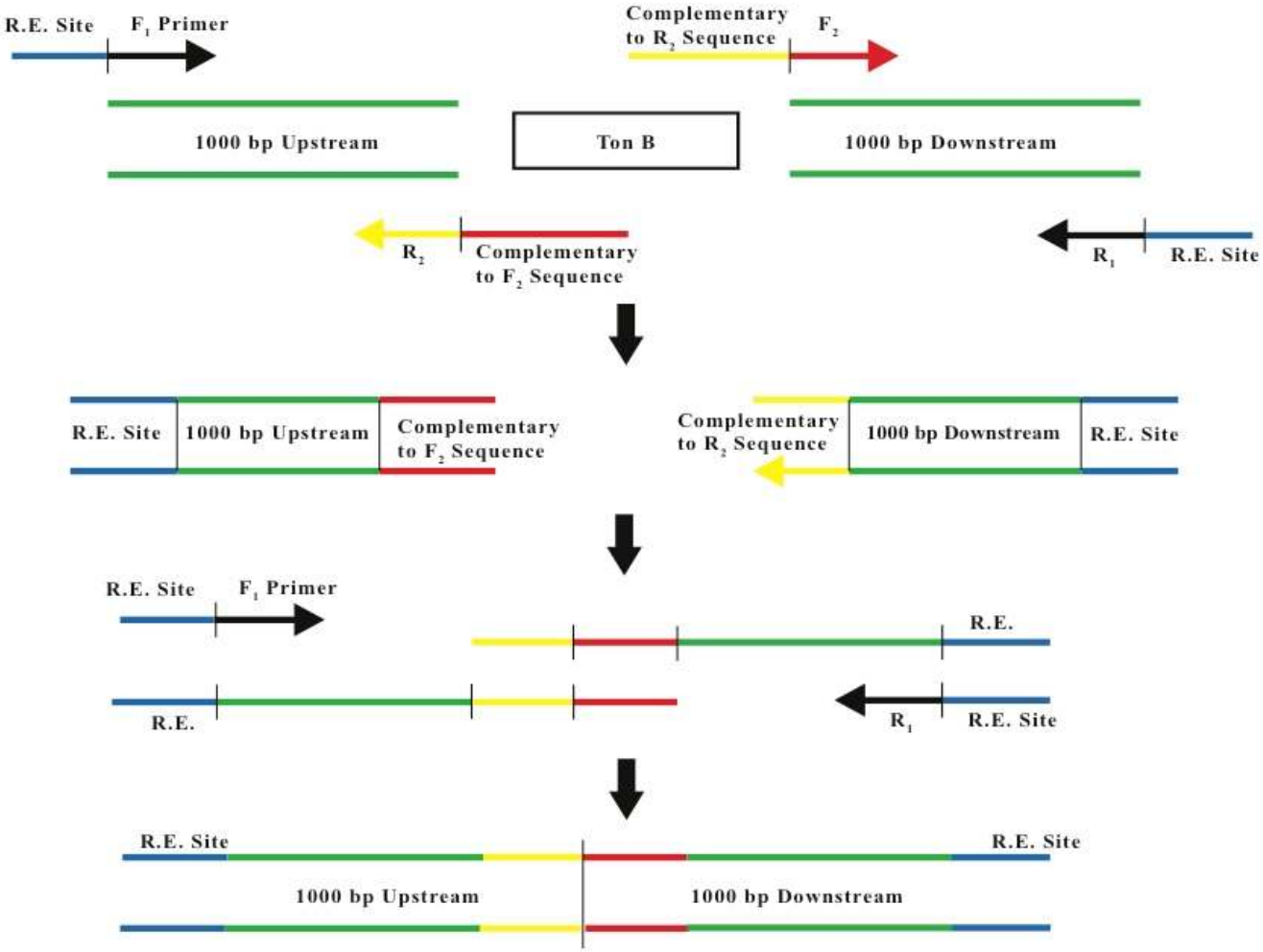

Figure 1. Splicing by overlap extension PCR. SOEing technique used to produce the construct capable of producing RLΔTB after undergoing homologous recombination

pEXTBS and Selecting for a Mutant

The pEX18 series of vectors were designed to facilitate homologous recombination to produce targeted mutagenesis (Hoang et al. 1998). The pEX series of vectors all contain a *sacB* counterselectable marker, pUC polylinker, and one of the

following antibiotic resistance markers: tetracycline (Tc), gentamicin (Gm), Ampicillin (Ap) (Hoang et al. 1998).

pEXTBS was produced by the integration of the SOE product into pEX18Gm and was subsequently transformed into SM10λpir by electroporation at 2.5kV/cm. Colonies were serial diluted onto LB_{Gm15} plates and screened via colony PCR. One colony that contained pEXTBS was used for biparental conjugation with *R. leguminosarum* ATCC 14479. Three mL of $YMB_{Na10Pg50}$ was inoculated with *R. leguminosarum* ATCC 14479 and placed onto a rotary shaker at 30° C for 48 hours. Three ml of LB_{Gm15} broth was inoculated with a SM10λpir colony harboring pEXTBS and was grown overnight at 37□°C on a rotary shaker. Cells were pelleted in 2.0 mL microcentrifuge tubes and washed 2 times with sterile saline (0.85% NaCl) to remove residual antibiotics. Cells were then resuspended in 200μL sterile saline. Two hundred μL of ATCC 14479 was combined with 100 μL of SM10λpir containing pEXTBS in a sterile microcentrifuge tube and vortexed. The mixture was pelleted, supernatant decanted, and resuspended in 30 μL of 10 mM $MgCl_2$. The mixture was then pipetted onto sterile 0.45 μm nitrocellulose filter discs resting on a warmed CR plate. The mixture was allowed to incubate for 24 hours at 30° C. After incubation, the mixture was washed off the disc by submersion in a microcentrifuge tube containing one mL of sterile saline and vortexed. The resulting mixture was serial diluted and plated on $CR_{Na10Gm10}$ plates and allowed to incubate for 72 hrs. Colonies were screened via colony PCR.

Merodiploids were described as colonies that had integrated pEXTBS into their genome and thus contained both the *tonB* and its flanking regions and the *tonB* SOE product. Once merodiploids were identified, they were transferred to 5 ml YMB_{Na10}. The

broth was allowed to incubate on a rotary shaker at 30° C for 48 hrs. After incubation, the culture was serial diluted and plated on CR media containing 10μg/ml nalidixic acid and 5% sucrose ($CR_{Na10Suc5}$) to counterselect against merodiploids that have not undergone homologous recombination and "kicked" out pEXTBS. Dilutions were allowed to incubate on $CR_{Na10Suc5}$ media at 30° C for 72 hours. Colonies were screened via colony PCR to identify potential mutants. A genomic extraction was then performed on each potential mutant. The genomic DNA of each potential mutant was used as template for PCR reactions using the primer set of F1-R1. These products were sent for sequencing to confirm mutants. Colonies confirmed to have an in-frame deletion of *tonB* were called *R. leguminosarum* RLΔTB.

High and Low Iron Growth for ATCC 14479 and RLΔTB

Growth curves were produced for wild type ATCC 14479 and RLΔTB in complex media, low iron, and high iron media. YMB_{Na10} was used as a complex media in 50 mL volume per 250 mL flask. Three flasks were inoculated with ATCC 14479 and 3 flasks were inoculated with RLΔTB at a 1:100 ratio. The flasks were shaken at 200 rpm at 30°C. Growth was measured every 4 hours at OD_{600}.

In low iron media, six 250 mL flasks were acid treated with nitric acid to scavenge contaminating iron. After acid treating, the flasks were washed 4 times with Millipore double distilled water. Six flasks were then filled with 50 mL of MMW_{Na10} media containing 0.25mM 2'2 dipyridyl and 0.25μM $FeCl_3$. Three flasks were inoculated with ATCC 14479 and 3 flasks were inoculated with RLΔTB at a 1:100 ratio from seed cultures of each strain grown in 3 ml of MMW_{Na10} containing 0.25 mM 2'2 dipyridyl and

5.0μM $FeCl_3$. The flasks were shaken at 200 rpm at 30°C. Growth was measured every 8 hours at OD_{600}.

In high iron media, six 250 mL flasks were acid treated with nitric acid to scavenge contaminating iron. After acid treating, the flasks were washed 4 times with Millipore double distilled water. Six flasks were then filled with 50 mL of MMW_{Na10} media supplemented with 0.25mM 2,2 dipyridyl and 5.0μM $FeCl_3$. Three flasks were inoculated with ATCC 14479 and 3 flasks were inoculated with RLΔTB at a 1:100 ratio from seed cultures grown in MMW_{Na10} supplemented with 0.25mM 2,2 dipyridyl and 5.0μM $FeCl_3$. The flasks were shaken at 200 rpm at 30°C. Growth was measured every 8 hours at OD_{600}.

pMRTBC

Rhizobia species are known to recognize incP (RP4 and RK2), incQ and pBB (Bordetella vectors) origins of replication (email correspondance with Dr. Phillip Poole, John Innes Centre, Norwich, UK). pMR10 is a broad-host range, low copy plasmid with an incP *ori* (RK2), pUC polylinker, and confers kanamycin resistance (Mohr and Roberts, unpublished). pMRTBC was produced by digestion of the F1-R1 fragment from pUCFR and its subsequent ligation into pMR10 (Fig. 2). As mentioned earlier the F1-R1 fragment contains the promoter region for the cotranscribed *orf1-tonB* genes along with *orf1* and *tonB*. Once the products were digested and ligated into pMR10, electroporated into DH5α, and screened via colony PCR, potential colonies were grown on LB_{Km50} plates. pMRTBC was mobilized into RLΔTB via triparental conjugation with the assistance of a helper strain harboring pRK2013.

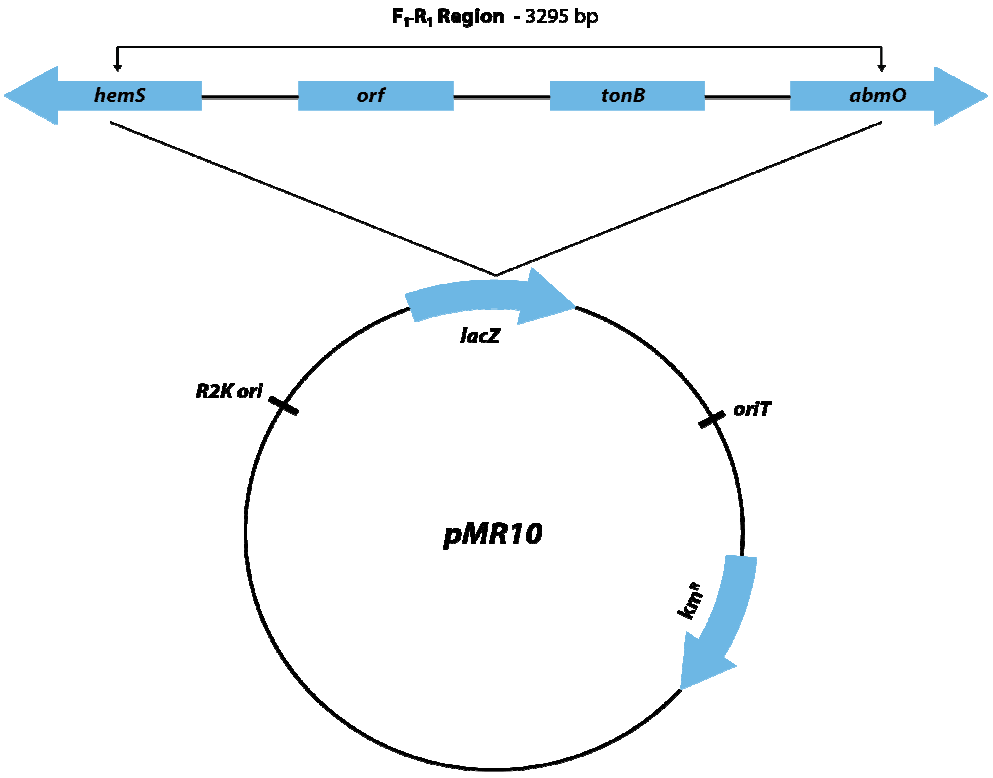

Figure 2. pMRTBC. Diagram of pMRTBC

Growth of RLΔTB harboring pMRTBC

Six 250mL flasks were acid treated with nitric acid to scavenge contaminating iron. After acid treating, the flasks were washed 4 times with Millipore double distilled water. Six flasks were then filled with 50 mL of MMW_{Na10} media containing 0.25 mM 2'2 dipyridyl and 0.25μM $FeCl_3$. Three flasks were inoculated with ATCC 14479 and 3 flasks containing 50μg/mL of kanamycin were inoculated with RLΔTB harboring pMRTBC at a 1:100 ratio from seed cultures of each strain grown in 3 ml of MMW_{Na10} ($MMW_{Na10Km50}$ for the mutant strain harboring pMRTBC) containing 0.25 mM 2'2 dipyridyl and 5.0μM $FeCl_3$. The flasks were shaken at 200 rpm at 30°C. Growth was measured every 8 hours at OD_{600}.

pMRTP1 and pMRTP2

pMRTP1 and pMRTP2 (Fig. 3) were generated using crossover PCR. The primer sets used to generate pMRTP1 were SOEF1-TBProR1 and TBProR1-SOER1, and the primer sets used to generate pMRTP2 were SOEF1-TBProR2 and TBProF2-SOER1. During the initial PCR reactions to produce the upstream and downstream products, pUCFR was used as template. The same procedure used to generate a SOE product was used to generate pMRTP1 and pMRTP2's products. Once the amplicon was produce via crossover PCR, it was digested and ligated into pMR10, electroporated into DH5α, screened via colony PCR, and colonies harboring the desired plasmid were grown on LB_{Km50}. The plasmid was sequenced to verify accuracy of the crossover PCR product.

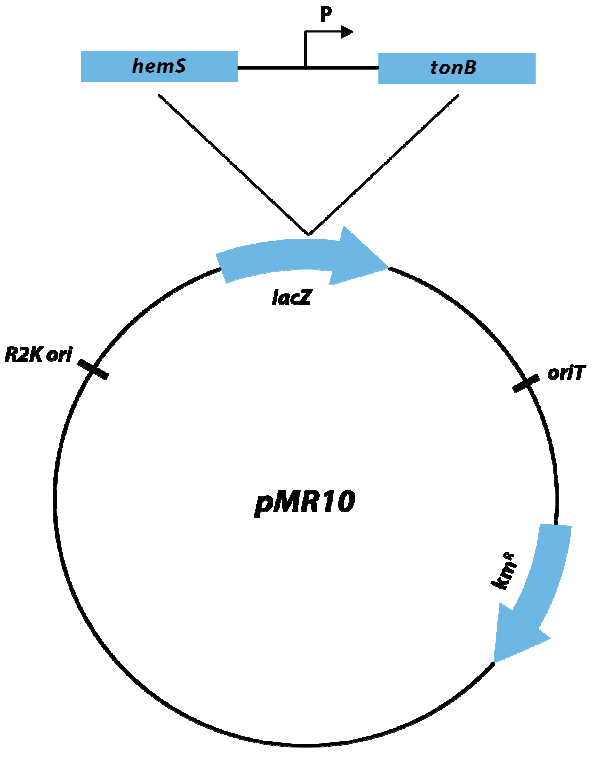

Figure 3. pMRTP1 and pMRTP2. Diagram of pMRTP1 and pMRPT2

pMRTP1 and pMRTP2 was mobilized into RLΔTB via triparental conjugation with the assistance of a helper strain harboring pRK2013.

Vicibactin Isolation and Purification

The isolation and purification of vicibactin followed the same procedure developed by Wright et al. (2013). Instead of using ATCC 14479 to produce vicibactin, ATCC 14479 harboring pBIO1187 was used for the overexpression of vicibactin (Fig. 4). pBIO1187 is an LAFR-1 based cosmid that contains the *vbs* operon and the ECF factor of the vicibactin operon, *rpoI* (Carter et al., 2002). pBIO1187 was introduced in wild type ATTC 14479 though triparental conjugation with helper plasmid pRK2013.

Batch cultures of ATCC 14479 harboring pBIO1187 were grown in 2.5L flat bottom flask containing 1L $MMW_{Na10Tc10}$ on a rotary shaker at 30° C for 96 hrs. Cells were then pelleted by centrifugation at 15,000 g at 4° C. Supernatant was collected in a 1L sterile, acid-washed bottle and acidified to pH of 2 using 6M HCl. Acidified supernatant was loaded into a 5 cm x 30 cm column containing 5cm x 8cm of

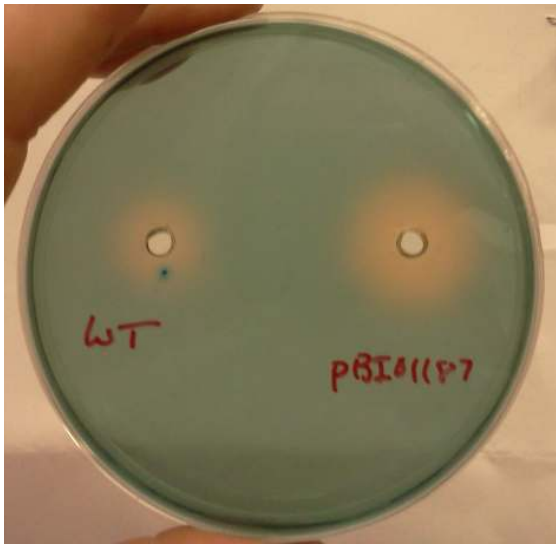

Figure 4. CAS assay of ATCC 14479 and ATCC 14479 harboring pBIO1187. CAS media plated with 50 □L of supernatant from ATCC 14479 (WT) and ATCC 14479 harboring pBIO1187.

AmberliteXAD-2 (approx. 160g) for crude purification. Prior to elution, 3 bed volumes of double distilled H_2O were run through the column to ensure all supernatant had passed through the XAD-2. Methanol was used to elute the column and samples were collected in 50 ml centrifuge tubes. CAS plates containing 1 cm (diameter) wells were then filled with 50 microliters of each sample (Fig. 5). Samples producing a halo and showing positive for the presence of vicibactin were combined and placed in a Buchi R-200 Rotovap to evaporate the methanol contained in the sample. During the evaporation process, the sample was subjected to a 30°C water bath for approximately 90 minutes. The remaining sample was then prepared for lyophilization on dry ice and placed at -80°C overnight. Samples were lyophilized overnight (or until dry) using Freeze Dryer Virtis Freezemobile on the campus of ETSU's Quillen Medical School. The dried samples were then resuspended in methanol and ran though approximately 45 cm of Sephadex LH-20 (25 g) packed into a 1.5 x 50 cm column. A total of 40 fractions were collected in 150 drop aliquots using a Bio-Rad Model 2110 fraction collector. A volume

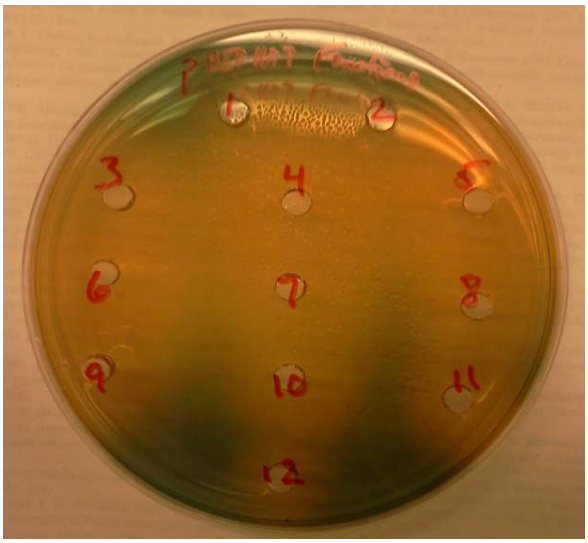

Figure 5. XAD-2 fractions of ATCC 14479 harboring pBIO1187. 50 µL was pipetted into each well.

of 5 μL of each fraction was spotted onto 20 x 20 TLC plates and run in a glass chamber using a solvent system containing n-butanol:acetic acid:ddH_2O (12:3:5). The plates were developed with 0.1 M $FeCl_3$ in 0.1 N HCl (Storey 2005). Fractions containing vicibactin were combined and dried using the Buchi R-200 Rotovap. Dried samples were stored at -20°C until further use.

Preparation of ^{55}Fe-vicibactin Complex

According to Wright et al., the purification of vicibactin leads to the isolation of an intact compound known as vicibactin C and two degradative compounds called vicibactin A and vicibactin B (Wright et al. 2013). Because there are differences within the structure of these 3 compounds, spectral absorbance can be used to identify the approximate amount of each molecule present. Using spectral absorbance analyses, it was determine that the intact vicibactin C's absorbance maximum was 450nm, which differed from vicibactin B and vicibactin A's absorbance maxima that was 465nm and 475nm, respectively (Wright et al. 2013). The molarity of the Sephadex LH-20 purified vicibactin was determined using the molar absorbance coefficient (1510 M^{-1} cm^{-1}, Carson et al. 1994). The vicibactin with a known concentration was used to prepare 2 solutions, ^{55}Fe-vicibactin and Fe-vicibactin. The stock solutions were made by mixing vicibactin at a 1:1 molar ratio with $^{55}FeCl_3$ purchased from Perkin Elmer and with $FeSO_4$. When different molar solutions of ^{55}Fe-vicibactin and Fe-vicibactin were produced, they were always made with a 0.1M MOPS solution. The mixtures were allowed to stir for an hour and stored at -20°C until further use.

Time-dependent Transport Assays

Transport experiments were conducted on ATCC 14479, RLΔTB, and RLΔTB harboring pMRTBC, pMRTP1, and pMRTP2. Seed cultures of each strain were grown in YMB broth with the appropriate antibiotic for 48 hours at 30°C. The seed cultures were then used to make a 1:100 inoculation of 50 mL MMW with the appropriate antibiotic in a 250 mL flask. Flask were shaken for at 30°C until OD_{600} of 0.4-0.6 (approx. 48 hrs.), at which point the cultures were transferred to two 50 mL conical tube (25 mL in each centrifuge tube) and pelleted at 10,500 x g for 1 hour at 4°C. The supernatant was then decanted and the cells were washed twice with a 0.85% NaCl solution to remove any excess exopolysaccharide that may interfere with the transport assays. After washing the cells were resuspended in 1 mL of MMW broth and transferred back into 50 mL of MMW containing the appropriate antibiotics. The cultures were then allowed to shake for approximately an hour before the transport assays.

The transport assays were performed in a 37°C water bath and went as follows. Ten milliliters of MMW was pipetted into a 50 mL test tube. Forty microliters of a 40μM ^{55}Fe-vicibactin were pipetted into the test tube containing 10 mL MMW to adjust the overall concentration to 80nM ^{55}Fe-vicibactin. One hundred microliters of culture was pipetted into a different 50 mL test tube. The test tube containing the 10 mL of 80nM MMW was then poured into the 50 mL test tube containing 100μL of culture and allowed to incubate for a period of time. At the end of an appropriate incubation period, 40μL of 0.25mM Fe-vicibactin was added to the mixture to stop the transport. After the Fe-vicibactin solution was added, the mixture was vacuum filtered through 0.45nm nitrocellulose filter and washed with a 0.9% lithium chloride solution, a chaotropic agent.

The incubation times used for the transport assays were 15 seconds, 30 seconds, 1 minute (min.), 2 min., 3 min., 4 min., 6 min., 8 min., 10 min., and 15 min. The experiments were performed in triplicate.

ATCC 14479 TonB Expression

pET17b (Novagen) is an expression vector designed for *E. coli*. pET17b contains an ampicillin resistance marker, a T7 RNA polymerase that is under lac promotion, and a multiple cloning site (MSC) that contains an upstream T7 promoter region. This configuration allows for an IPTG induced transcription of cloned sequences. When IPTG is added to media containing *E. coli* harboring pET17b, the lac promoter is induced to transcribe the T7 RNA polymerase. The T7 RNA polymerase then binds specifically to the T7 promoter region to transcribe the desired sequence. Upstream of the MCS lies a T7 tag that is attached to the N-terminus of proteins expressed by this vector. The T7 tag allows for rapid identification with T7 antibody.

In this study we attempted to produce a 200 amino acid C-terminal product containing a portion of the proline-rich region and a 120 amino acid C-terminal product. The primer pairs were as follows: TonBPrCt and Rev17b to produce pET200 and TonBCT and Rev17b to produce pET120. The pUCFR was used as template in the PCR mixture to produce the desired amplicons. After digestion and ligation, each ligation mixture was used to transform BL21 via electroporated at 2.5kv/cm. Transformation mixtures were serial diluted and plated onto $LB_{Amp100.}$ Colonies containing the desired plasmid were identified by colony PCR. Candidates identified via colony PCR were prepped and sequenced.

Expression of pET120 and pET200

Once pET120 and pET200 were identified, three 3 ml seed cultures were started. These broths were incubated overnight at 37□°C on a rotary shaker. Seed cultures were used to inoculate three, 250 mL flasks containing 50 mL of LB_{Amp100}. When cultures reached 0.4-0.6 at OD_{600}, 1mM IPTG was added to each culture to induce protein expression. Each culture was allowed to continue shaking for 3 hours. After 3 hours of incubation, 200 µL of each culture was pipetted into a 2.0 mL microcentrifuge tube and pelleted. Approximately 15 to twenty 200 µl samples of each culture were pelleted, supernatant discarded, and stored at -20°C until further use.

SDS-PAGE Visualization of C-terminal Domains

Frozen samples of BL21 harboring pET120, pET200, pETFL were resuspended in 50□l of Tris-HCl and 50□l 2x loading dye and boiled at 95°C for 10 minutes. After boiling, the mixtures were allowed to cool and subsequently centrifuged for 5 minutes at 13,200 rpm to pellet cellular debris. The samples were then loaded into a 10% polyacrylamide gel made with the recipe in Table 1. The gels were run at 200V/60 amps for approximately 1 hour. Gels were then removed and stained overnight with Coomassie Blue.

Western Blotting of C-terminal Domains

Proteins expressed using the pET17b vector system was visualized by SDS-PAGE were then transferred to PVDF. SDS-PAGE gels, fiber pads, and Whatman #1 filter paper were first placed in transfer buffer for 20 minutes. Millipore® polyvinylidene floride

(PVDF) membranes were first cut to the desired size and then soaked in methanol for 5 minutes. After soaking, the PVDF membrane was washed with ddH_2O and placed in transfer buffer for 10 minutes. The "sandwich" was then created in the following order: fiber pad, filter paper, SDS-PAGE gel, PVDF membrane, filter paper, and fiber pad. The "sandwich" was secured in the support grid, placed in the electrophoretic apparatus, and submerged in ice-cold transfer buffer. The western blot was allowed to run overnight at 4˚ C over-night at 25-30 volts.

The PVDF membrane containing transferred protein fragments was first subjected to Novagen® T7•Tag® Monoclonal Antibody and later to Novagen® T7•Tag® Antibody HRP conjugate using the protocol supplied by Novagen®. Following the addition of primary and secondary antibody, the gels were stained with either 4-chloro-1-naphthol or Thermo Scientific™ Pierce™ ECL 2 Western Blotting Substrate.

CHAPTER 3

RESULTS AND DISCUSSION

ATCC 14479 TonB Sequence

The *tonb* sequence amplified in ATCC 14479 was estimated to be 457 amino acids long and was very similar in length to the putative *tonb* of *R. leguminosarum* 8401, which was estimated at 454 amino acids long (Johnston et al. 2001). The 2 putative protein sequences also shared 96.5% similarity when aligned by Expasy SIM (Fig. 6).

```
96.5% identity in 457 residues overlap; Score: 2241.0; Gap frequency: 0.7%

Rl14479,       1 MNPGGHGQTVRGMQMAISAKSRSRQVLIGEPDADGSLNDNNMHPGHELSDLRNVQRQPAG
Rl8401,        1 MNPGGHGQTVRGIQMAISAKSRSRQVLIGEPDADGSLNDNNMHPGHELSDLRNVQRQPAG
                 ************ **********************************************

Rl14479,      61 EAVVHYARFAQIPSFPDHPEAEPIASVPAPPIDAAVEKQEDERKPMRRRMALTCIGSFFF
Rl8401,       61 EAVVHYARFAQIPAFPDHPEAEPIASVPAPPIDAAVEKQEDEKKPVRRRVALTCIGSFFF
                 ************* ************************** ** *** **********

Rl14479,     121 HAGLVVALSVFMPTPPNEAIEDAGEAVSVVVYGDSDVDQTSAGDPELERQPEQVASEEVE
Rl8401,      121 HAGLLVALSVFMPTPPNEAIEDAGEAVSVVVYGDSDVDQTSAGDPELQRQPEQVASEEVE
                 **** ***************************************** ************

Rl14479,     181 PDTVQSEEATELAATTVPPEQAQPVETETMETVQSVQEVTRVSPEAVVAAEPEVLVSESP
Rl8401,      181 PDTVQSEEATELNATTVPPEQAQPVETETMETVQSVQEVTRVSPEAVVAAEPEVLVSESP
                 ************ **********************************************

Rl14479,     241 AESFVAQPMATAVPEQPMPDIAQATVPEEVVPTAVQPTAVPPEEVKPVETVEVSPEPEDK
Rl8401,      241 AESFVAQPMATAVPEQPMPDIAQATVPEEVVPTAVQPTAVPPEEVK---TVEVSPEPEDK
                 ********************************************   ***********

Rl14479,     301 PKPKAKKPKPVEKKQSQKRAKPAGGKEGSDREDSTRGMVNGQDGPQTDGTSTTTGGTDGM
Rl8401,      298 PKPKAEKPKPVEKKQSQKRAKPAGGKEGSDREDSTRGTVNGQHGPQTDGTSTTTGGTDGM
                 ***** ****************************** **** *****************

Rl14479,     361 GSAAVANYPGKIQKRIRRAVRVPDEYKNKSGGMTVRIQLTINGTGRVASVSVARSSGIAE
Rl8401,      358 GSAAVANYPGKIQKRIRRAVRVPDEYKNKSDGMTVRIQLTINGTGRVASVSVARSSGIAE
                 **************************** ****************************

Rl14479,     421 LDKAVLDGVRRAAPFPPLPSEWGKPSWTFAQEVQVTR
Rl8401,      418 LDKAVLDGVRRAAPFPPLPSEWGKPSWTFTQEVQVTR
                 *************************** *******
```

Figure 6. ATCC 14479 and 8401 TonB Expasy SIM amino acid alignment. *R. leguminosarum* ATCC 14479 and *R. leguminosarum* 8401 are respresented as Rl14479 and Rl8401, respectively.

Next, the amino acid sequences of the TonB proteins' of *E. coli* and ATCC 14479 were aligned using ClustalW (Fig. 7). The sequences were also aligned using T-Coffee

(Fig. 8), which assessed physical-chemical similarity, and Dompred (Fig. 9), which predicted secondary structure. The conserved domains and "SHLS" motif that was putatively identified in *R. leguminosarum* 8401 was also identified within the TonB of ATCC 14479 when aligned with the *E. coli* TonB.

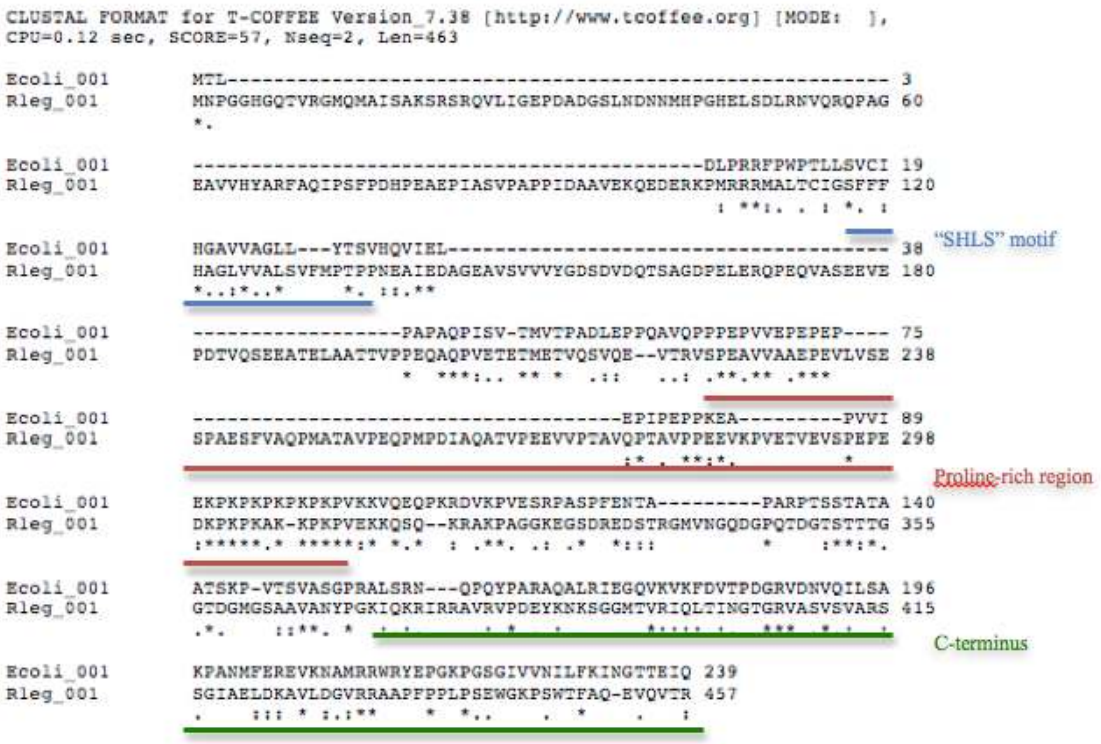

Figure 7. The TonB amino acid sequences of ATCC 14479 and *E. coli* aligned by ClustalW . *R. leguminosarum* ATCC 14479 and *E. coli* are respresented as Rleg_001 and Ecoli_001, respectively.

N-terminal Domain

In *E. coli*, the N-terminus has been determined to contain residues 1-32 with the "SHLS" motif being comprised of S16, H20, L28, and S31 (Larsen et al. 1999). According to Dompred (Fig. 9), the "SHLS" motif of *E. coli* forms an α-helix, which is a part of transmembrane domains. In ATCC 14479, Dompred also predicts an α-helix in

the stretch of sequence that ClustalW had aligned with the *E. coli* "SHLS" region. Within this sequence, a putative "SHLS" region comprised of residues S117, H120, L124 or L128, and potentially S129 is found. The residues L124, L125, and S129 may or may not contribute to the formation of the motif, because it has been previously shown through substitution experimentation that the only essential residue within the motif of *E. coli* is H20 (Swayne and Postle 2011). In ATCC 14479, this histidine is present at residue 120.

The translational start of the *R. leguminosarum* putative TonB is not known but was predicted based on the typical ATG start codon and the location of a potential ribosomal binding site (Wexler et al. 2001). If this translational start is correct then ATCC 14479's TonB contains an N-terminus that is 135 amino acids long with a potential cytoplasmic tail approximately 116 amino acids long. This domain would be much larger than the 32 amino acids of the *E. coli* N-terminus. However, a BLAST of the 115 amino acids that comprise the putative extended N-terminus shared no sequence similarity or conserved motifs to any known genes.

Proline-rich Region

In *E. coli*, the proline-rich region has been identified to comprise residues 66-102 (Khursigara et al. 2005). According to the ClustalW alignment (Fig. 7), the proline-rich region of ATCC 14479 appears to comprise residues 225-310. An analysis of the amino acid distribution of this region showed that within the *E. coli* TonB the most prevalent residues were proline at 40.5%, glutamate at 21.6%, lysine at 18.9%, and valine at 10.8%. In ATCC 14479, the most prominent residues were proline at 20.7%, valine at 18.4%, glutamate at 16.1%, and alanine at 13.8% (data not shown). After analysis by amino acid

distribution, a hypothesis can be made as to why the proline-rich region of ATCC 14479 might need to be 49 amino acids longer that than of *E. coli*. The additional length of the proline-rich region in ATCC 14479 may be necessary due to the lower percentage of proline residues. As was previously mentioned, the proline-rich region has been hypothesize to provide a rigid extension that allows the C-terminal domain to span the periplasmic space in order to contact the TBDT residing in the outer membrane. These additional residues could compensate for a less rigid central region and allow the protein to overcome the distance of the periplasmic space. It is possible that *E. coli* has evolved a more efficient central region than has *R. leguminosarum*.

However, if the central portion of the 2 TonBs is determined to begin after the "SHLS" motif and conclude at just prior to the first α-helices of the C-terminal domain, then the percentages of amino acids within the regions change. The most prominent to least prominent residues within the 102 amino acid sequence of *E. coli* becomes proline at 19.6%, valine at 16.5%, glutamate at 14.7%, alanine at 12.7%, and lysine at 10.8%, and ATCC 14479 percentages from residue 132-323 becomes glutamate at 16.8%, valine at 15.7%, proline at 15.2%, and alanine at 11.5%. Although proline continues to occur most often, the percentage of proline between the *E. coli* and ATCC 14479 become more similar (19.6% to 15.2%, respectively) and the most common residues (proline, valine, alanine, and glutamate) within this region are the same for both proteins.

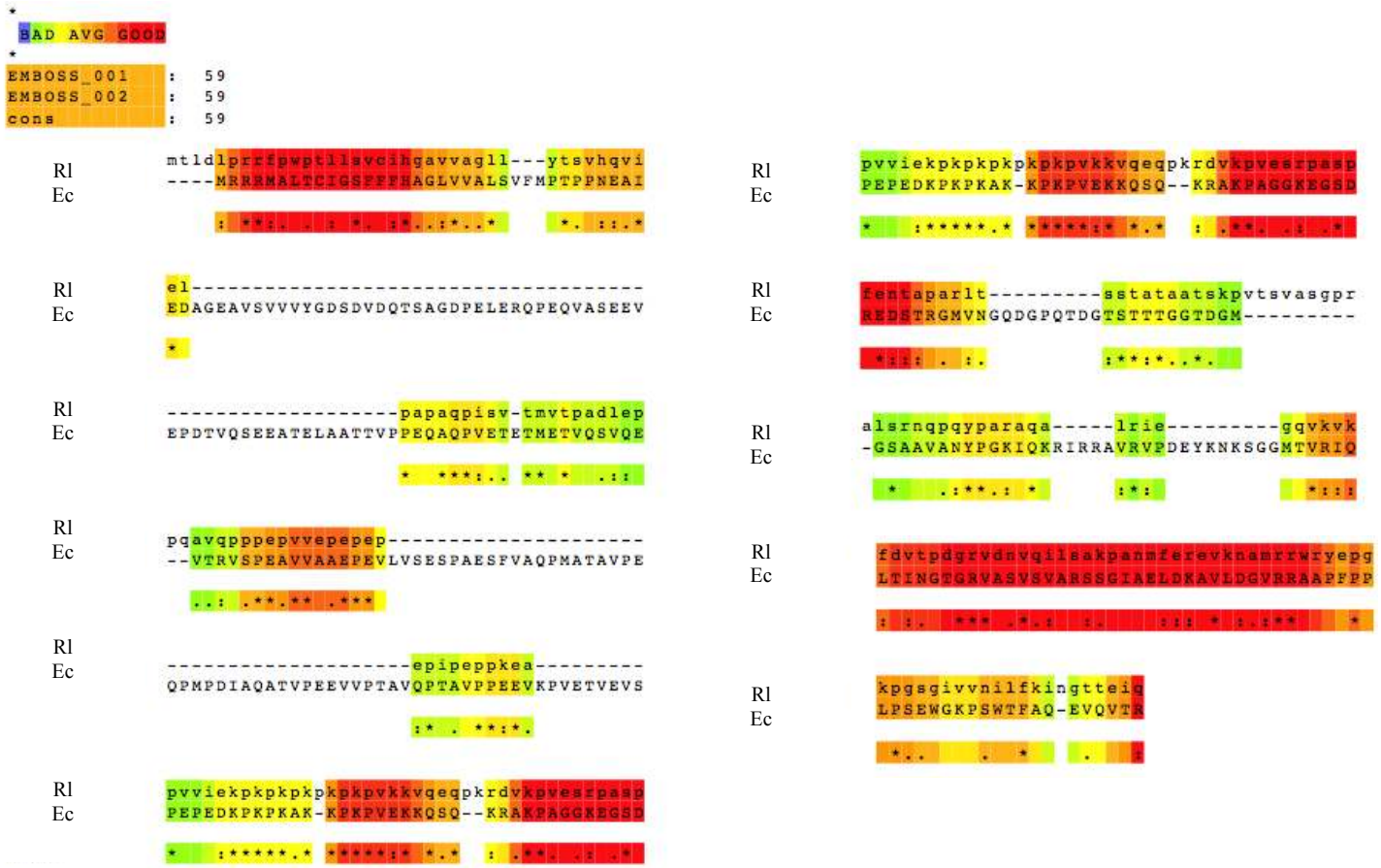

Figure 8. TonB amino acid sequences of ATCC 14479 and *E. coli* TonB aligned using T-Coffee. *R. leguminosarum* ATCC 14479 and *E. coli* are respresented as RL and EC, respectively.

ATCC 14479 *E. coli*

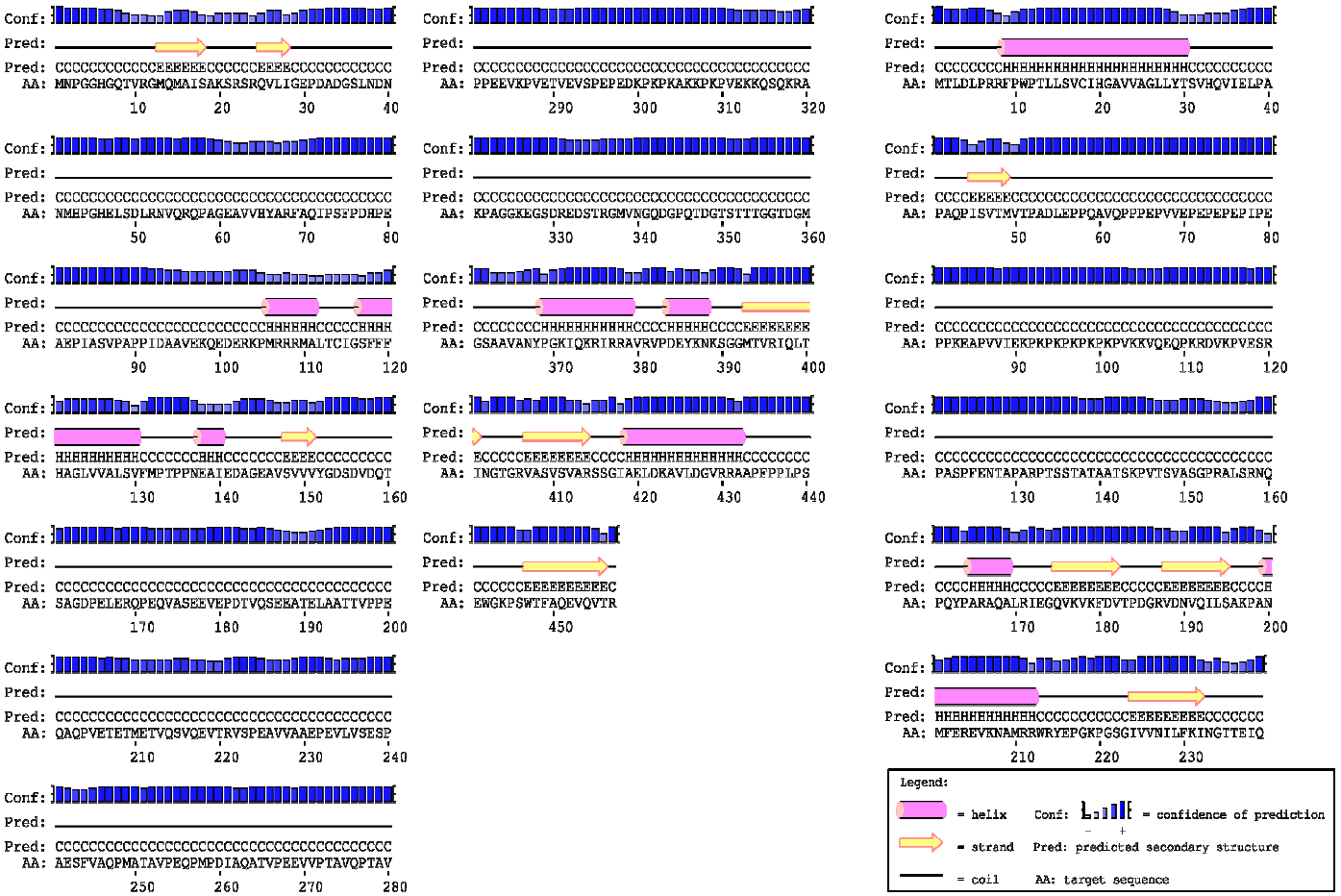

Figure 9. TonB amino acid sequences of ATCC 14479 and *E. coli* aligned using Dompred. Dompred was designed by the UCL Department of Computer Science Bioinformatics Group.

C-terminal Domain

The C-terminal domain of *E. coli* has been determined to comprise residues 155-239. A crystal structure (Fig. 10) of TonB, residues 158-235, in complex with FhuA have determined that the C-terminal of TonB consist of 3 β strands and 2 α–helices (Pawalek et al. 2006). β_1 and β_2 are comprised of residues 174 and 197 and β_3 comprised of 223 to 231. A short and long α helix, α_1 and α_2, being comprised of residues 165-170 and 203-210, respectively (Pawalek et al. 2006). The crystal structure also elucidated the interaction between the C-terminual β_3 strand of TonB and the TonB box, which results in the formation of an interprotein β-sheet (Pawalek et al. 2006).

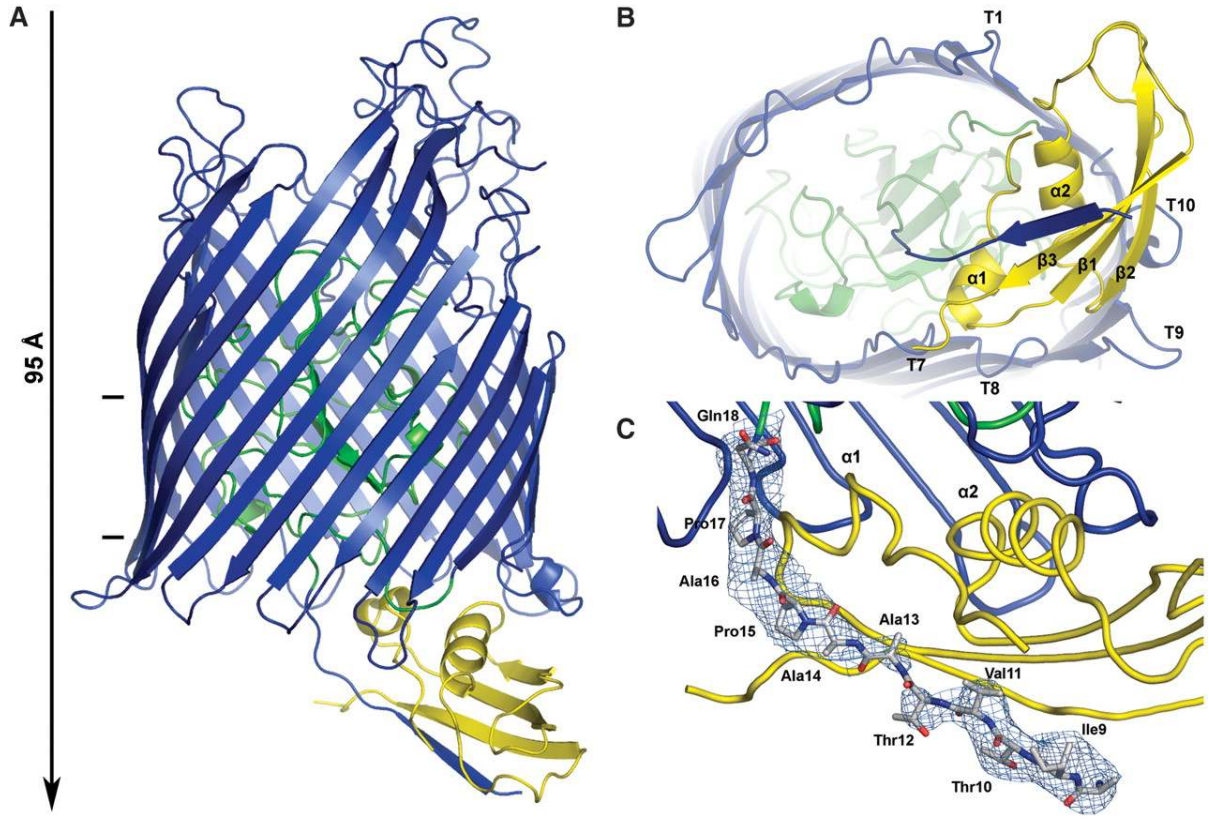

Figure 10. The TonB C-terminus and FhuA co-crystallization. The image shows the FhuA barrel (blue) and plug domains (green) along with TonB C-terminus (yellow) interacting with FhuA's N-terminally located "TonB box". Diagram provided courtesy of Pawalek (Pawalek et al. 2006).

Figure 9 shows the predicted secondary structures found within the amino acid sequences of ATCC 14479's TonB and *E. coli*'s TonB. Because the crystal structures and the Dompred prediction of *E. coli*'s TonB not only shared the same 2 α-helices and 3 β strands that comprise the C-terminus but also the identical sequences in which these secondary structures were determined to exist, a certain confidence can be given to the accuracy of the Dompred predictions. In ATCC 14479, the C-terminus was predicted to contain not only the same secondary structures but also the same ordering of secondary structures as the sequence progressed. The short and long α helix, α_1 and α_2, being comprised of residues 384-388 and 419-432, respectively, and the 3 β strands, β_1, β_2, and β_3, were comprised of residues 393-401, 406-414, and 447-456, respectively. The only difference between the C-terminals of ATCC 14479 and *E. coli* was the prediction of an additional α-helix at residues 369-379 within the ATCC 14479 TonB. When analyzing results generated by T-Coffee's amino acid sequence alignment (Fig. 8), the physical-chemical similarity between the proteins seem to be the greatest between residues 181-218 of *E. coli* and residues 400-437 of ATCC 14479. This region contains a portion of the β_1 strand, all of the β_2 strand, and the long α helix. The short α helix and the β_3 strand seem to share fewer functional characteristics.

Identification of RLΔTB

Upon the completion of the selection and counter-selection protocols for mutant identification, colony PCR was used in the indirect identification of a *tonB* mutant. Figure 11 contains the products amplified from PCR reactions containing the F1-R1

primer set with wild type, merodiploid, and mutant genomic DNA used as template. The amount of the PCR mixture loaded into each lane varied to produce the best visual representation of the results. Once a mutant colony was identified, primers external to the F1-R1 primer sets were designed to insure the correct location of the recombination event. The primer pair hybridized to an upstream *hemS* sequence and a downstream putative antibiotic monooxygenase gene, *abmO* and was thus named HemS and AbmO. To confirm the isolation of a mutant with an in-frame deletion of *tonB*, the product amplified by the HemS-AbmO primer pair was sequenced (sequence found in Appendix II). ATCC 14479 with an in-frame deletion of *tonB* was named RLΔTB.

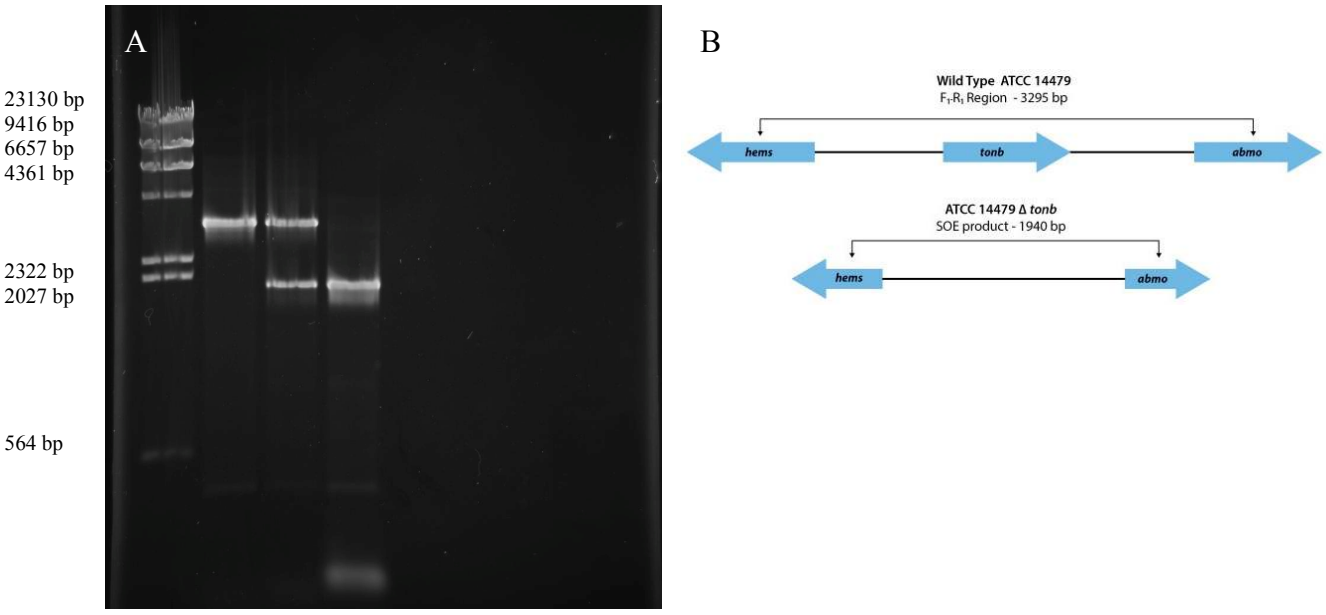

Figure 11. Agarose gel showing F1-R1 products using wild type, merodiploid, and mutant genomic DNA. (A) 1% TBE agarose gel of ATCC 14479 (lane 2), merodiploid (lane 3), and RLΔTB (lane 4) F1-R1 products. (B) Illustration of ATCC 14479 (wildtype) and RLΔTB F1-R1 regions.

Growth Curves of ATCC 14479 versus RLΔTB

Once RLΔTB had been identified, 3 growth curves were generated in order to compare mutant and ATCC 14479 growths.

Complex Media

Figure 12 represents the first growth curve generated and shows that in complex media there is no statistical difference between the growth curves of RLΔTB and wild type ATCC 14479. Because the composition of the media may vary due to the inclusion of yeast extract, an exact quantity of each nutrient cannot be calculated. Therefore, an exact explanation as to why RLΔTB and ATCC14479 generated similar growth curves cannot be given. However, we hypothesize that the growth curves were similar because the complex media contained an iron concentration high enough to allow diffusion to meet this nutritional requirement, and the production and acquisition of iron through siderophore-mediated transport was not necessary.

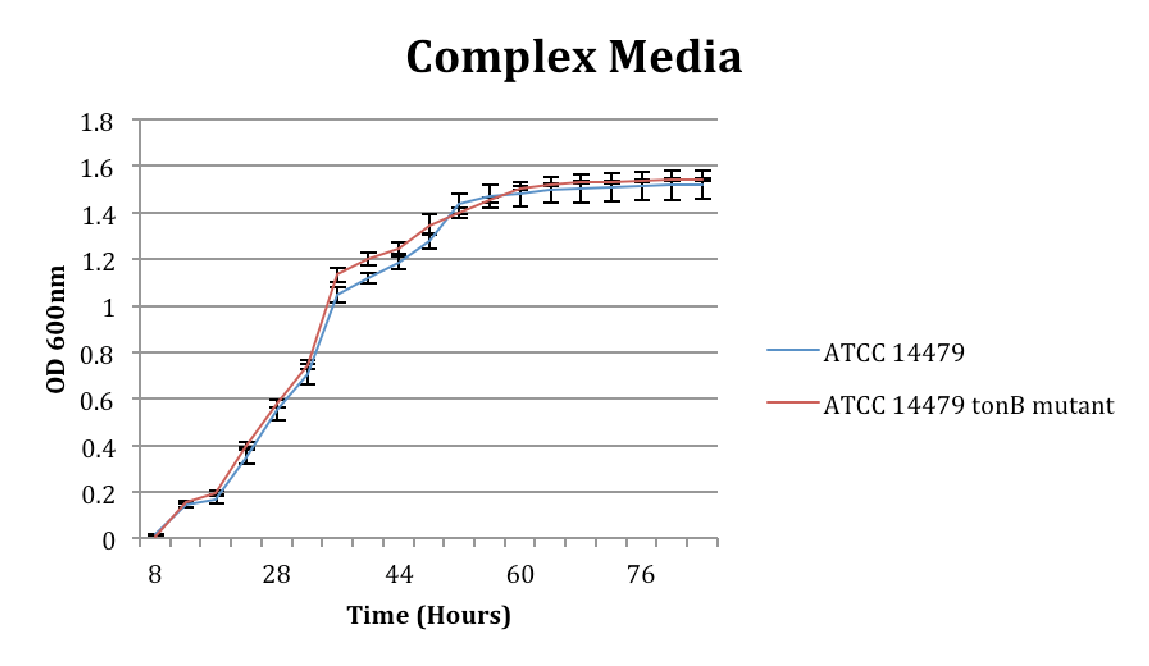

Figure 12. Growth curve of ATCC 14479 versus RLΔTB in complex media. Strains were grown in YMB_{Na10}.

High and Low Iron Media

To further test this hypothesis, an attempt to control the nutritional ingredients within the media was made. A minimal media, MMW, was used to control the iron concentration. Furthermore, the glassware was acid-washed and 2,2 dipyridyl was added to a final concentration of 0.25 mM. The supplementation of 0.25mM 2,2 dipyridyl would hypothetically bind any contaminating soluble iron, while allowing a variable amount of ferric iron, $FeCl_3$, to be added to the media. Figure 13 was generated by comparing the growth curves of RLΔTB and ATCC 14479 when grown in $MMW_{Na10Pg50}$ supplemented with 0.25 mM 2,2 dipyridyl and 5.0μM $FeCl_3$, and was consider high iron media. Figure 14 was generated by comparing the growth curves of RLΔTB and ATCC 14479 when grown in $MMW_{Na10Pg50}$ supplemented with 0.25 mM 2,2 dipyridyl and

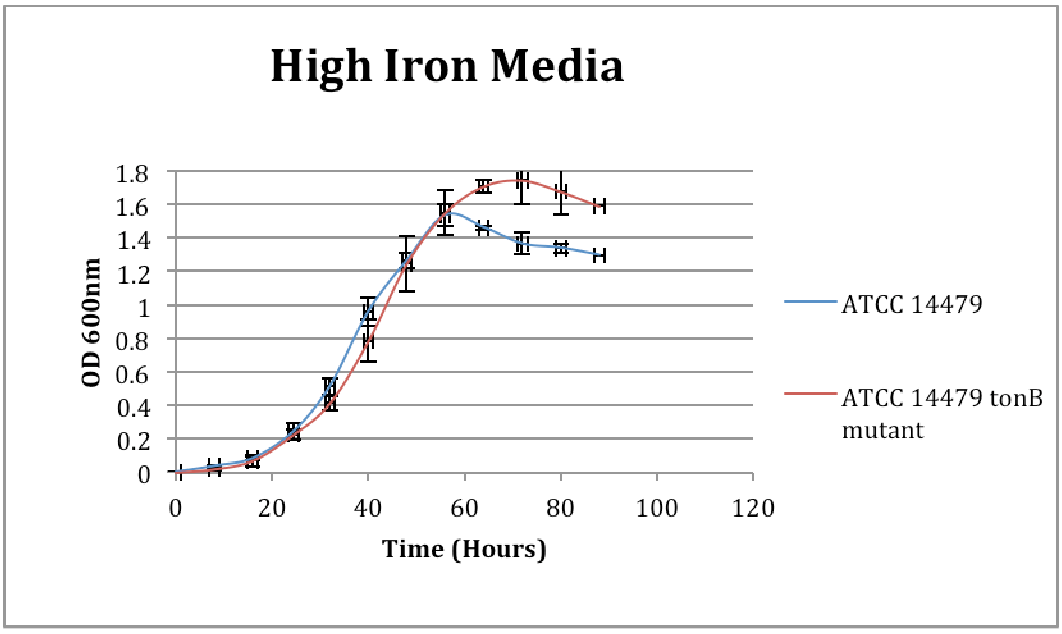

Figure 13. Growth curve of ATCC 14479 versus RLΔTB in High Iron Media. Strains were grown in $MMW_{Na10Pg50}$ supplemented with 0.25mM 2,2 dipyridyl and 5.0mM $FeCl_3$.

0.25μM $FeCl_3$, and was considered low iron media. Statistically, growth between wild type and mutant in high iron media showed no difference. However, when grown in low iron media, the mutant's growth was regressed compared to wild type growth. These results provide indirect evidence that when all other variable are relatively the same, iron played a determining factor between the growth curves generated by wild type ATCC 14479 and mutant RLΔTB. Also, these experiment provide indirect evidence that the stretch of sequence eliminated from the genome of RLΔTB plays a role in the transport of ferric iron.

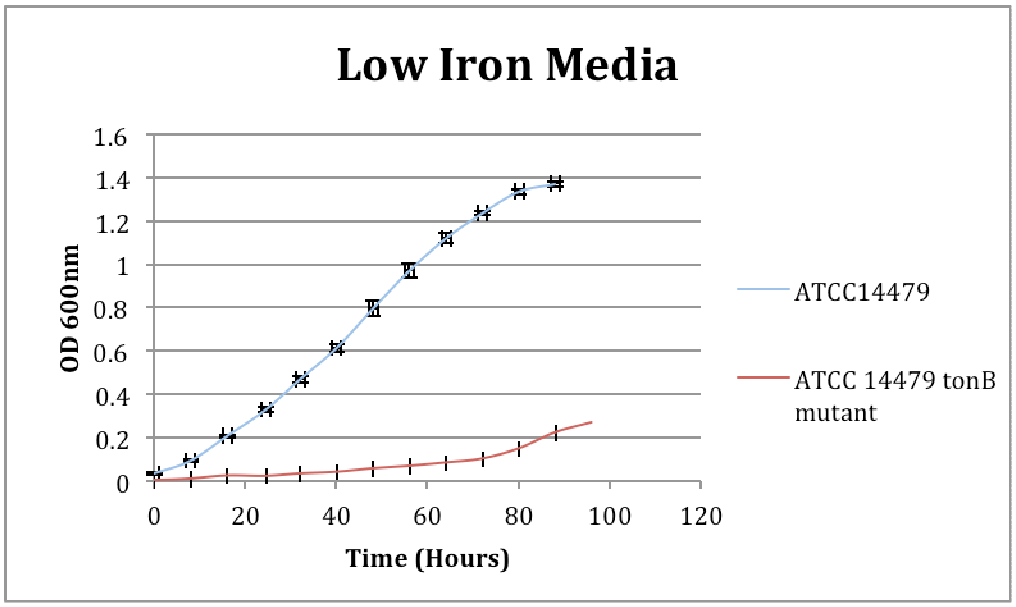

Figure 14. Growth curve of ATCC 14479 versus RLΔTB in Low Iron Media. Strains were grown in $MMW_{Na10Pg50}$ supplemented with 0.25mM 2,2 dipyridyl and 0.25mM $FeCl_3$.

Transport Assay of ATCC 14479 and RLΔTB

Transport assays using ^{55}Fe-vicibactin were performed to provide direct evidence that the regressed growth seen in RLΔTB was not only associated with the differing concentration of iron between the high and low iron minimal medias but more

specifically a result of RLΔTB inability to transport iron intracellularly due to the in-frame deletion of a putative *tonB* gene. Figure 15 shows that the phenotype exhibited by RLΔTB when grown in low iron minimal media was in fact due to the strains inability to transport iron, a growth-dependent factor, intracellularly. It is hypothesized that the baseline amount of ^{55}Fe detected in the mutant strain is a result of the binding of the ^{55}Fe-vicibactin complex to an outer membrane transporter.

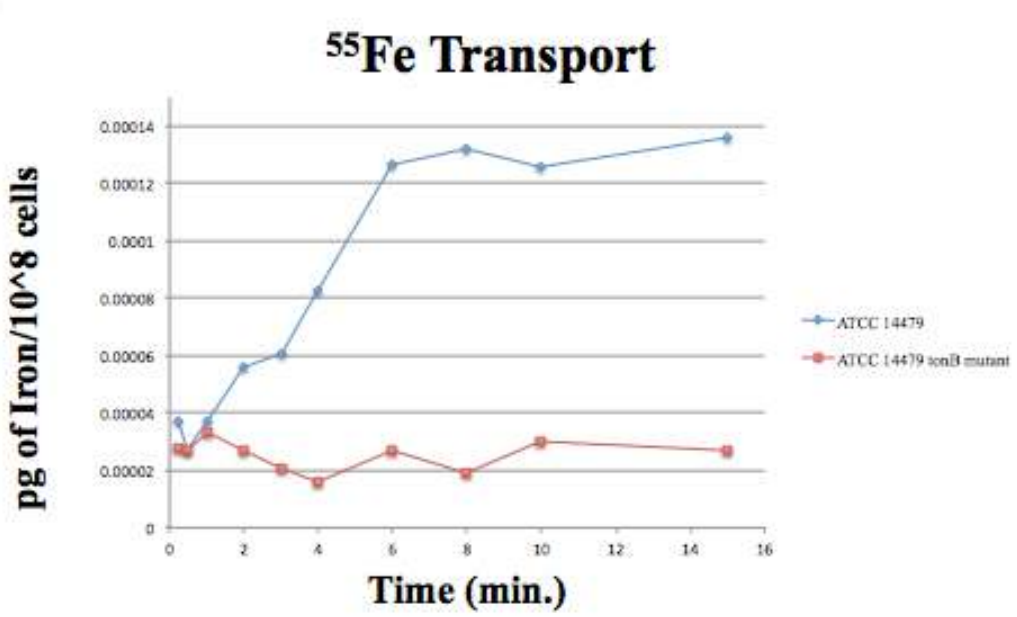

Figure 15. ^{55}Fe-Vicibactin Time-Dependent Transport assay of ATCC 14479 and RLΔTB. Wild type ATCC 14479 (blue) and mutant RLΔTB (red).

RLΔTB harboring pMRTBC

Because the translational start of the putative TonB has yet to be determined, the immediate implementation of the conventional method of complementation may have proven difficult. The choice to clone many sequences with different translational start sites into an inducible expression vector may introduce too many variables if complementation were unsuccessful. For example, if a sequence failed to complement the mutant, would it be due to the incorrect concentration of the inducer molecule or the

incorrect sequence (in this case, too short of a sequence may exclude a critical or unknown conserved motif in the N-terminus necessary for proper functioning) and how to decipher between the two? It is for this reason an alternative approach was chosen to identify the sequence of the *tonB* gene. Once identified, the ORF for *tonB* could then be used to restore wild type phenotype.

It was previously found in *R. leguminosarum* 8401 that upstream of *tonB* lies an ORF, *orf1*, with no known function (Wexler et al. 2002). Subsequently, primer extensions determined that *orf1-tonB* are co-transcribed and constitute an operon with the promoter region falling in the intergenic region between *hmuS* and *orf1* (Wexler et al. 2002). Analysis of the sequence of the F1-R1 fragment of ATCC 14479 showed that *orf1-tonB* operon was located just downstream of the *hmuS* and the noncoding region between the two operons shared 98.6% similarity (Fig. 16).

With this information, it was decided that the entire F1-R1 fragment could be cloned into a low copy expression vector and the natural promoter would be able to regulate transcription. If the F1-R1 fragment was capable of restoring phenotype without induction then it could be assumed, but not confirmed without RT-PCR or Northern blotting, that the functional TonB was expressed. From here, portions of this F1-R1 fragment could be deleted piece by piece to until the minimum *tonB* sequence capable of complementation was identified.

The F1-R1 fragment was ligated into a low copy vector, pMR10, forming pMRTBC and was mobilized by triparental conjugation into RLΔTB (Fig. 17). Consideration also went into choosing the low copy vector, pMR10 as oppose to one of the higher copy pBBR1MCS vectors for these experiments. It was believed that because

the native promoter was being used as opposed to an inducible promoter, a natural expression would occur.

```
  98.6% identity in 352 nt overlap (1-352:1-350); score: 1701 E(10000):
2.6e-134

              10        20        30        40        50        60
8401   GTCCTACCCTACTTATTGAGAATGAGCTTGCCCTGACGGGTGATCTTCAGGCGATAGACC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14479  GTCCTACCCTACTTATTGAGAATGAGCTTGCCCTGACGGGTGATCTTCAGGCGATAGACC
              10        20        30        40        50        60

              70        80        90       100       110       120
8401   AAGCCGTCGTGTCTAATCATGATCTCGTTCGTGCCGCGGAAGAGATCCGCGCTTTCGACG
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14479  AAGCCGTCGTGTCTAATCATGATCTCGTTCGTGCCGCGGAAGAGATCCGCGCTTTCGACG
              70        80        90       100       110       120

             130       140       150       160       170       180
8401   ATCCGGTGCTGCGCCGCAGGCTCGCTCTGCAGCGGCACGTGCTTAAAGTTATCTGGCTTT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
14479  ATCCGGTGCTGCGCCGCAGGCTCGCTCTGCAGCGGCACGTGCTTAAAGTTATCTGGCTTT
             130       140       150       160       170       180

             190       200       210       220       230       240
8401   TCAACCATCATTTCAGTTGGCAATTCCGCGAGACCGGTGATCAGGGTCCCCGGTTGTGAT
       :::::::::::::::::::::::::::::: :::::::::::::::::::::::::::  ::
14479  TCAACCATCATTTCAGTTGGCAATTCCGTGAGACCGGTGATCAGGGTCCCCGGTTG--AT
             190       200       210       220       230

             250       260       270       280       290       300
8401   GACAATTATCTTGACTTTCTTACTCATAGTTTTTTAAAGACGCAATAGGAGACTAATATA
       :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::: :
14479  GACAATTATCTTGACTTTCTTACTCATAGTTTTTTAAAGACGCAATAGGAGACTAATACA
      240       250       260       270       280       290

             310       320       330       340       350
8401   GTCAAGTTTTTGAGATTGCTGATATGAAGCCGCCTGGCGGCGGTTCGCGACA
       :::::::::::::::::: :::::::::::::::::::::::::::::::::
14479  GTCAAGTTTTTGAGATTACTGATATGAAGCCGCCTGGCGGCGGTTCGCGACA
       300       310       320       330       340       350
```

Figure 16. Nucleotide sequence of the intergenic regions of 8401 and ATCC 14479 aligned by lalign. ATCC 14479 and 8401 are represented by 14479 and 8401, respectively.

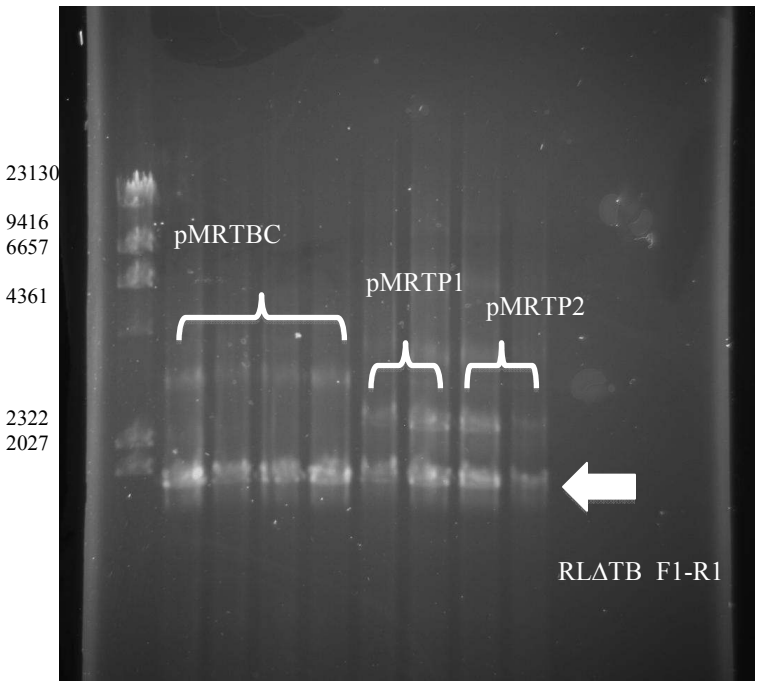

Figure 17. Agarose gel of F1-R1 products from RLΔTB harboring pMRTBC, pMRTP1, and pMRTP2. Lanes 2-5 RLΔTB harboring pMRTBC, lanes 6 and 7 RLΔTB harboring pMRTP1, and lanes 8 and 9 RLΔTB harboring pMRTP2.

Growth Curve and Transport Assay of RLΔTB harboring pMRTBC

Figure 18 shows the growth curves generated by ATCC 14479, RLΔTB harboring pMRTBC, and RLΔTB grown in low iron media. The results indicate that the mutant harboring pMRTBC showed growth patterns similar to ATCC 14479 and not that of the RLΔTB. Figure 19 shows a comparison between ATCC 14479, RLΔTB, and RLΔTB harboring pMRTBC in their ability to transport ^{55}Fe-vicibactin. Along with the restoration of growth, pMRTBC also restored vicibactin mediated iron transport in RLΔTB. Additionally, since the vector was not induced by IPTG, a restoration of wild type phenotype lends credibility to the primer extension experiments (Wexler et al. 2002) performed and provided evidence for the location of a promoter region just upstream of the *orf1-tonB* operon.

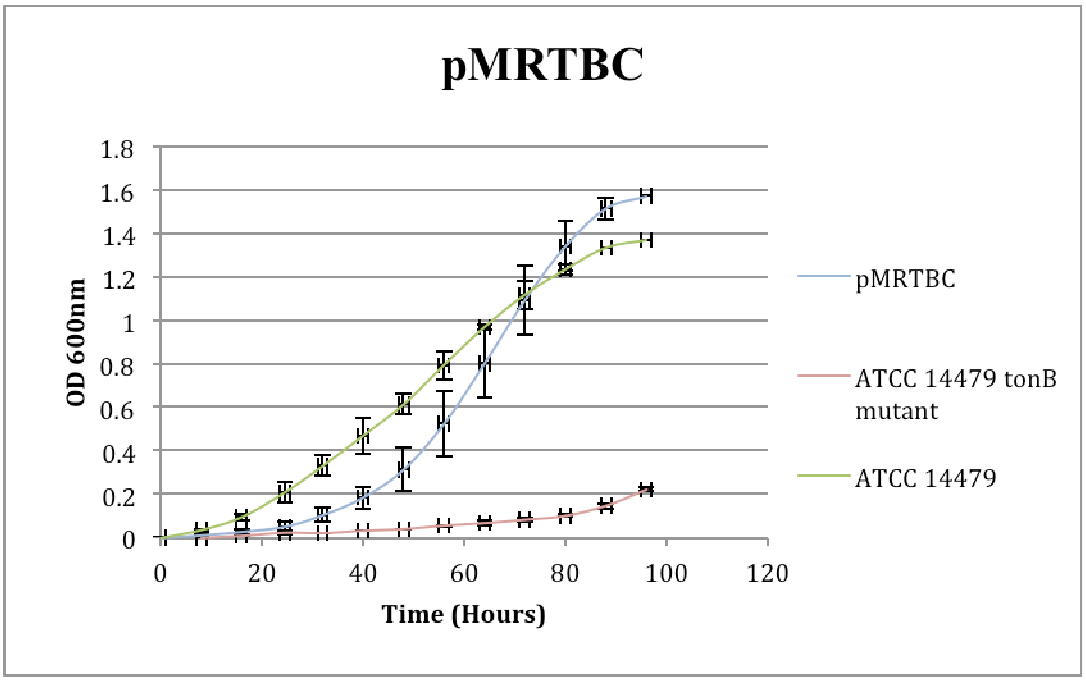

Figure 18. Growth Curve of wild type ATCC 14479, RLΔTB harboring pMRTBC, and mutant RLΔTB in low iron media. Strains were grown in $MMW_{Na10Pg50}$ supplemented with 0.25mM 2,2 dipyridyl and 0.25mM $FeCl_3$.

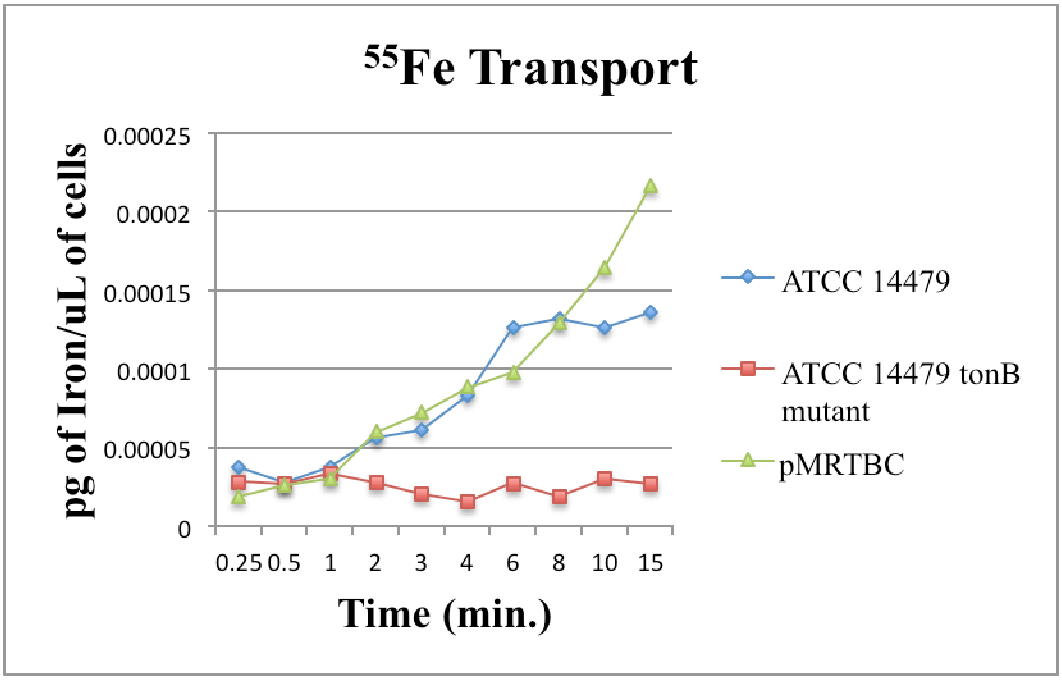

Figure 19. [55] Fe-Vicibactin Time-Dependent Transport assay of wild type ATCC 14479, RLΔTB harboring pMRTBC, and mutant RLΔTB. Wild type ATCC 14479 (blue), RLΔTB harboring pMRTBC (green), and mutant RLΔTB (red).

pMRTP1 and pMRTP2

pMRTBC provided evidence that within the F1-R1 region, a promoter to the *orf1-tonB* operon existed along with the ORF sequence necessary to restore wild type phenotype of both growth and 55Fe-vicibactin transport. The next step was to try and identify the sequence that was responsible for the restoration of wild type phenotype so that an accurate complementation construct could be designed.

As was mentioned earlier, the translational start codon of the TonB of *R. leguminosarum* is unknown. However, it has been predicted based upon a potential ribosomal binding site (RBS) and the more commonly used ATG start codon (Wexler et al. 2001). This potential start codon would produce an N-terminal domain that is 129 amino acids long, which is well extended compared to the 32 amino acid long N-terminus of the TonB of *E. coli.* In *R. leguminosarum,* the distance from the start codon to S116 at the beginning of the "SHLS" motif is 115 amino acids. In other gram-negative species (Fig. 20), the distance from the start codon to the "SHLS" motif is typically 12-15 amino acids. Also, analysis of the ATCC 14479 F1-R1 sequence revealed another potential RBS located within the N-terminus that would considerably shorten the predicted length of the domain and produce a protein that had a distance of 10 amino acids between the

Species	**"SHLS" Motif**	**# of Amino Acids from Start to Motif**
E. coli	SVCIHGAVVAGLLYTS	15
S. enterica	SVGIHGAVVAGLLYTS	13
K. pneumoniae	SVAIHFAVVAGLLYTS	15
V. cholerae	SLAFHALLLITTDEAQ	12
ATCC 14479	SFFFHAGLVVALSVFM	unknown

Figure 20. Distance from translational start to "SHLS" motif in different Gram-negative bacteria.

translational start codon and the "SHLS" motif (Fig. 21). This distance would be more similar to other Gram-negative TonB proteins. With this information, pMRTP1 and pMRTP2 were produced (Fig. 17). pMRTP1 contains the promoter region along with the original RBS and translational start codon predicted by Wexler (Wexler et al. 2001) and would produce a TonB that is 457 amino acids long. pMRTP2 also contains the promoter region, but instead of the original RBS and start codon, pMRTP2 contains the aforementioned alternative translational start codon that would produce a TonB of 353 amino acids long. It is important to note that within these constructs the upstream *orf1* has been spliced from the operon. This would provide evidence that the *orf1* sequence provides no role in the restoration of wild type phenotype and/or in the regulation or function of the putative TonB.

Figure 22 and Figure 23 show the results of the studies comparing RLΔTB harboring pMRTP1 and pMRTP2, respectively, in their ability to transport ^{55}Fe-vicibactin. In both experiments, RLΔTB harboring pMRTP1or pMRTP2 showed elevated levels of ^{55}Fe when compared to RLΔTB but decreased levels of transport when compared to wild type ATCC 14479.

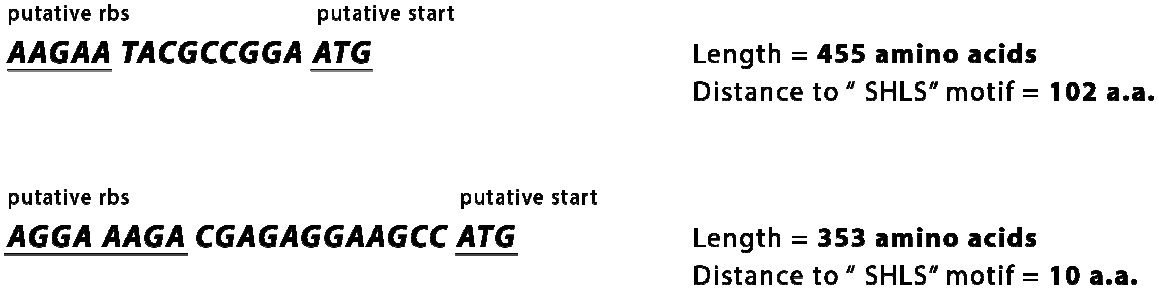

Figure 21. Putative RBS and translational start sites of pMRTP1 and pMRTP2.

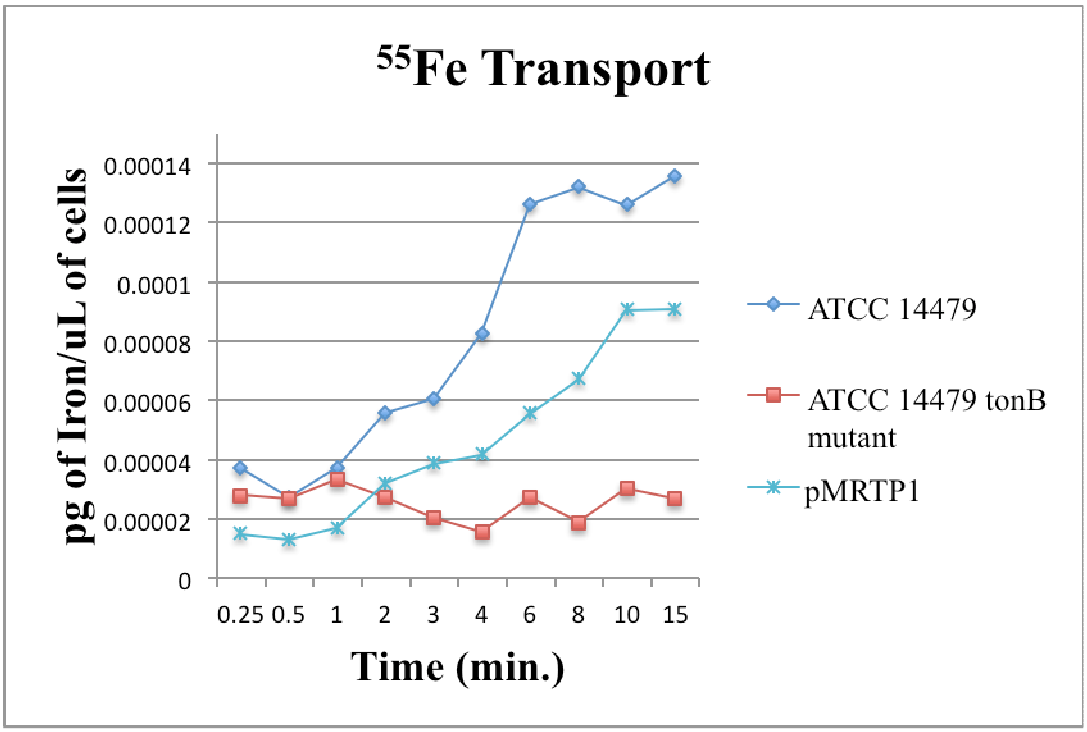

Figure 22. ^{55}Fe-Vicibactin Time-Dependent Transport assay of wild type ATCC 14479, RLΔTB harboring pMRTP1, and mutant RLΔTB. Wild type ATCC 14479 (blue), RLΔTB harboring pMRTP1 (teal), and mutant RLΔTB (red).

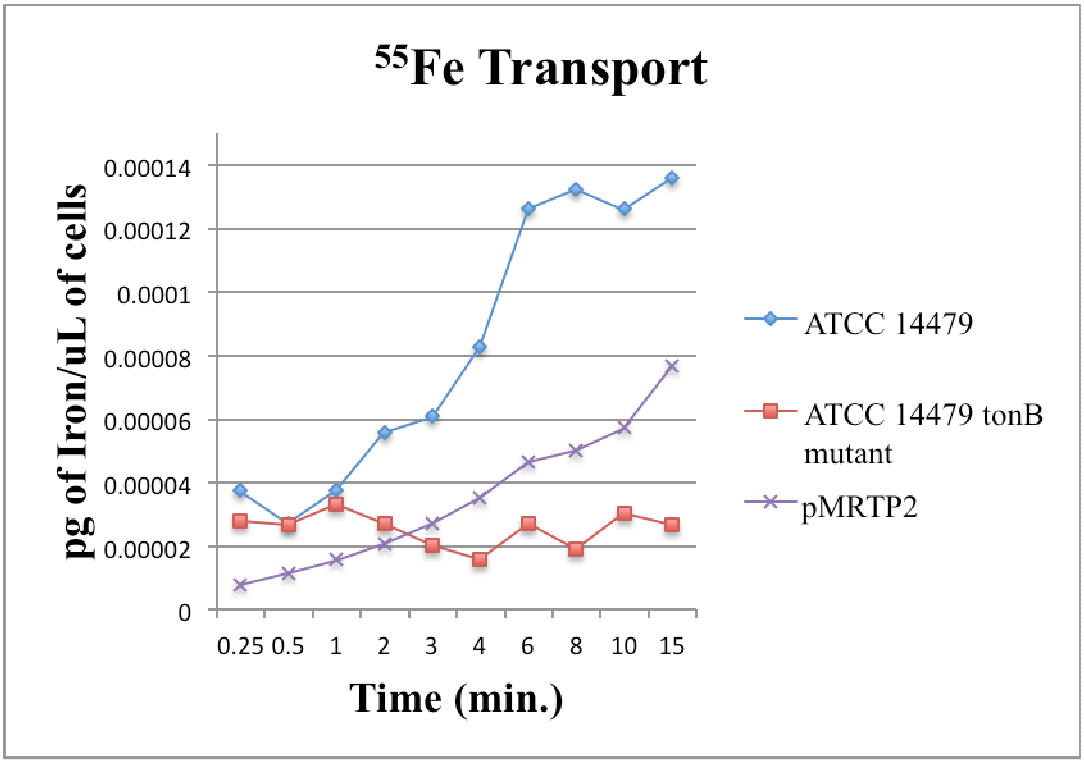

Figure 23. ^{55}Fe-Vicibactin Time-Dependent Transport assay of wild type ATCC 14479, RLΔTB harboring pMRTP2, and mutant RLΔTB. Wild type ATCC 14479 (blue), RLΔTB harboring pMRTP2 (purple), and mutant RLΔTB (red).

Expression of TonB

As was mentioned in the introduction, many groups have attempted to perform structural analysis of the *E. coli* TonB C-terminus by crystallization and NMR spectrometry. These attempts have often led to mixed results. Therefore, an attempt to express the C-terminus of the TonB of ATCC 14479 was undertaken. Two different sized C-terminal fragments, 200 and 120 amino acids long, were cloned into the *E. coli* expression vector pet17b and became pET200 and pET120, respectively. The 200 amino acid long C-terminal fragment contained residues M338-R457. The 120 amino acid long C-terminal fragment contained residues M258-R457. Because the C-terminal fragments were being expressed in pET17b, they include a T7 tag attached to the N-terminal end for ease of identification. The 2 lengths were chosen for 2 reasons. First, if these fragments could be overexpressed, purified, and able to undergo structural analysis by NMR or crystallization, there should be consistency within the 2 structures. Second, the difference between the 120 and 200 amino acid long fragments is that the 200 amino acid long fragment contains a portion of the proline-rich region. This was included so that if these fragments could be structurally analyzed, the proline-rich region should remain rigid and separate from the 3-dimensional configuration the C-terminal domain takes on. Therefore, lending credibility to the folding of the C-terminal domain. Expression of a 200 amino acid long and 120 amino acid long C-terminal TonB fragments from pET200 and pET120, respectively, from whole cell lysates were visualized first on SDS-PAGE (Fig. 24). Once large quantities of protein were seen at molecular weights equivalent to the estimated weights of the individual protein fragments, the proteins were mobilized onto PVDF for western blot analysis. Both proteins were visualized when stained with 4-

chloro-1-naphthol (Fig. 25) and only the 120 amino acids long fragment was present when stained with Pierce ECL Western Blot Substrate (Fig. 26).

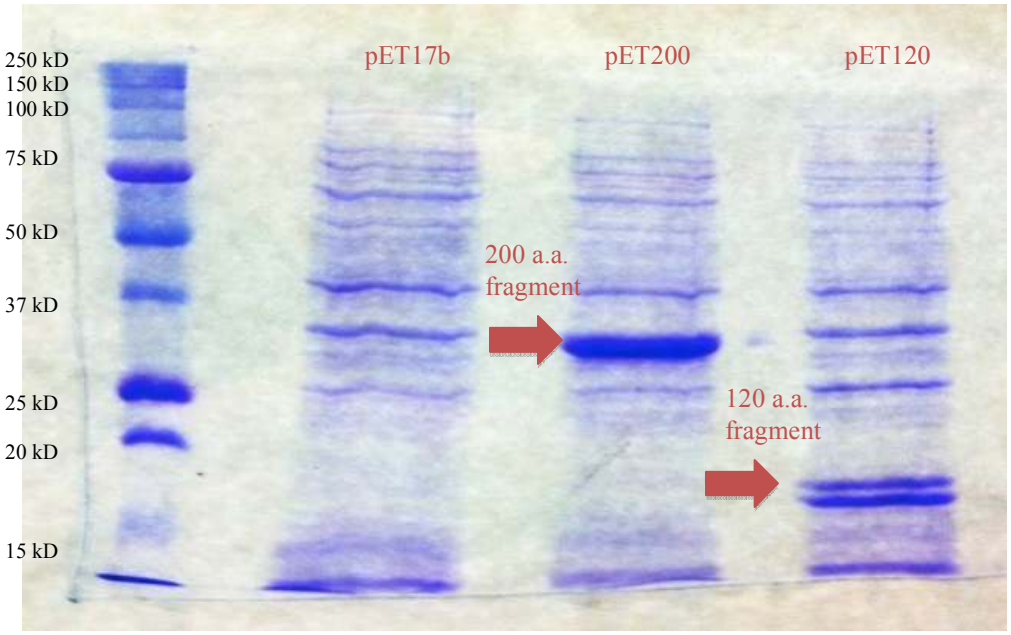

Figure 24. SDS-PAGE of overexpressed C-terminal TonB fragements of pET120 and pET200. Lane 2 pET17b control, Lane 3 pET200, and Lane 4 pET120.

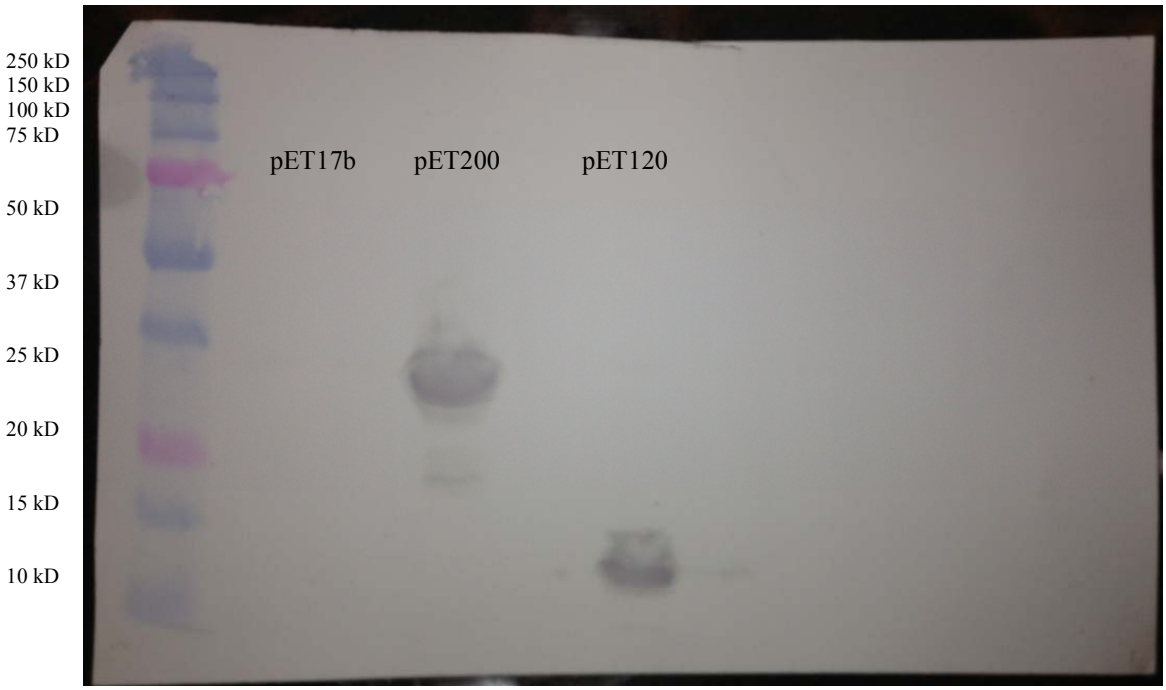

Figure 25. Western Blot of overexpressed C-terminal TonB fragements of pET120 and pET200 using 4-chloro-1-naphthol. Lane 2 pET17b control, Lane 3 pET200, and Lane 4 pET120.

Figure 26. Western Blot of overexpressed C-terminal TonB fragements of pET120 and pET200 using Pierce ECL Western Blot Substrate. Lane 2 pET17b control, Lane 3 pET200, and Lane 4 pET120.

CHAPTER 4

CONCLUSION

In this work a putative *tonB* was identified with an amino acid sequence that was 96.5% similar to that of *R. leguminosarum* 8401 (Wexler et al. 2001). When compared to the well-documented TonB of *E. coli*, ATCC 14479's TonB shared many similarities within the N-terminus, proline-rich region, and C-terminus when analyzed for predicted secondary structure and physical-chemical properties. However, there were differences between the 2 proteins. Depending upon the correct translational start site for ATCC 14479's TonB, there may be a well-extended N-terminus that contains no conserved motifs. The putative protein also possesses a lengthened and less proline-rich central domain that may rely upon the additional amino acids to compensate for the lack of rigidity. Lastly, the C-terminal of ATCC 14479's TonB is predicted to contain an identical ordering of secondary structure as that of *E. coli's* TonB. However, in the third beta strand, which is most important for interacting with the "TonB box" of TBDTs, there is almost no sequence or physical-chemical similarity. Furthermore, a quick T-Coffee comparison (data not shown) of FhuA and a putative hydroxamate TBDT found in *R. leguminosarum* showed that little to no sequence or physical-chemical similarity existed in the "TonB box" of these proteins. So, it may be that the 2 proteins from 2 different species are interacting with different "TonB boxes" with different characteristics and thus the third beta strand of the 2 TonBs must have different properties to engage in these interactions.

Next, a strain, RLΔTB, with an in-frame knockout of *tonB* was engineered. RLΔTB showed similar growth patterns as wild type when grown in complex media and

minimal media supplemented with 5.0μM $FeCl_3$ but showed a repressed growth compared to wild type when grown in the same minimal media supplemented with 0.25μM $FeCl_3$. Furthermore, the mutant was unable to transport ^{55}Fe-vicibactin when compared to wild type. Thus continuing to provide evidence that not only does this putative gene has sequence and proposed structural similarity to other known TonB proteins, but it also has a role in iron acquisition via vicibactin mediated iron transport.

An attempt to determine the ORF of ATCC's TonB was then performed. It had been previously determined through primer extensions that the promoter of the *orf1-tonB* operon lay in the intergenic region between *hemS* and *orf1* (Wexler et al. 2002). Because the ATCC 14479 F1R1 fragment contained this intergenic region and the *orf1-tonB* operon, it was hypothesized that the integration of the F1R1 fragment into a low copy vector, pMRTBC, could regulate itself due to the presence of the native promoter and restore wildtype phenotype to RLΔTB without induction. RLΔTB harboring pMRTBC not only restored wild type growth to the mutant in low iron media but also restored its ability to transport ^{55}Fe-vicibactin. These finding supported the notion that the promoter was present and was capable of regulating TonB expression without induction.

After the pMRTBC experiments, pMRTP1 was constructed. pMRTP1 contained the intergenic region, no *orf1*, and the original RBS and translational start site (Wexler et al. 2002) integrated into pMR10. Next, pMRTP1 was constructed, which contained the intergenic region, no *orf1*, and an alternative RBS and translational start site also integrated into pMR10. Both partially restored RLΔTB's ability to transport ^{55}Fe-vicibactin to wild type levels. However, an explanation could be that pMRTP1 contains the original RBS and translational start site along with the alternative RBS and

translational start site, while pMRTP2 contains only the alternative RBS and translational start site. With the commonality between the 2 being the alternative RBS and translational start site, a vector should be construct containing only the alternative translational start site to stop codon to see if wild type transport is restored.

Lastly, 2 C-terminal TonB fragments, 200 and 120 amino acids long, were overexpressed in pET200 and pET120, respectively. However, there was inconsistency in visualizing the 200 amino acids fragment via western blot. On one occasion a band can be seen at the predicted molecular weight using chemoluminescence and not on other occasions. It would be worthwhile to strain the PVDF after transfer to make sure that the protein was initially present in the SDS-PAGE gel and if so, was it transferred. The 120 amino acids long fragment was always visualized via western blot. However, it can only be confirmed as the C-terminal fragment via N-terminal sequencing or TonB C-terminal specific antibodies.

It was assumed that the promoter region existed in front of the *orf1-tonB* operon based on the work of Wexler et al. (2002). It could just be as likely that the *lac* promter of pMR10 is leaky and is the actual cause of transcription. To check for a leaky promoter, the F1-R1 sequence could be cloned into pMR10 backwards and if the phenotype restored, it could be assumed the promoter was present. Second, RT-PCR has not be conducted to determine whether the sequence is being transcribed or more importantly, the expression of the putative TonB proteins from pMRTBC, pMRTP1, and pMRTP2 have not been confirmed via western blot with TonB specific antibodies.

Moving forward on this project, it is necessary to produce a complementation construct that contains the correct *tonB* sequence and is induced to restore wild type

phenotype. It is also critical to draw antibodies specific for TonB and confirmed its expression is consistent with the restoration of wild type phenotype. Also, with these antibodies TonB can be identified in ATCC 14479, N-terminal sequenced, and the true *tonB* sequence determined. It is also critically important to use these antibodies to identify TonB's location in the inner membrane where it complexes with ExbB-ExbD and its activity with outer membrane TBDTs by using the expressed C-terminal fragment. These experiments would show that this putative protein is a TonB and not just another protein that with a role in siderophore mediated iron transport. It would also be interesting to determine whether the areas of the proline-rich region that are different and show no similarity to the proline-rich region of *E. coli* have a role in the proton translocation pathway. Once a TBDT of ATCC 14479 is identified, it would be worthwhile to attempt a co-crystallization with the expressed C-terminal domain of ATCC 14479 to see if a interprotein β-sheet is formed between the TonB 'box' and the third β-strand of the ATCC TonB C-terminal, or if another binding mechanism exist. There are many questions left for examination and hopefully a sturdy foundation has been laid for future endeavors.

REFERENCES

Agarwal S, Sebastian S, Szmigielski B, Rice PA, Genco CA. 2008. Expression of the gonococcal global regulatory protein fur and genes encompassing the fur and iron regulon during in vitro and in vivo infection in women. Journal of Bacteriology 190(9): 3129-3139.

Andrews SC, Robinson AK, Rodriguez-Quinones F. 2003. Bacterial iron homeostasis. FEMS MIcrobiology Review 27(2-3): 215-237.

Anzaldi LL, Skaar EP. 2010. Overcoming the heme paradox: Heme toxicity and tolerance in bacterial pathogens. Infectious Immunology 78(12): 4977-4989.

Athavale SS, Petrov AS, Hsiao C, Watkins D, Prichett CD, Gossett JJ, Lie L, Bowman JC, O'Neill E, Bernier CR, et al. 2012. RNA folding and catalysis mediated by iron (II). PLoS One *5*(e38024).

Baker KR, Postle K. 2013. Mutations in *Escherichia coli* ExbB transmembrane domains identify scaffolding and signal transduction functions and exclude participation in a proton pathway. Journal of Bacteriology 195(12): 2898-2911.

Bertini I, Gray HB, Stiefel EI, Valentine JS. 2007. Biological inorganic chemistry: Structure and reactivity. Journal of Chemistry Education 84(9): 1432.

Brenneisen P, Went J, Lkozu O, Wlaschek M, Briviba K, Krieg T, Sies H, Scharffetter-Kochanek K. 1998. Central role of Ferrous/Ferric iron in the ultraviolet B irradiation-mediated signaling pathway leading to increased interstitial collagenase (matrix-degrading metalloprotease (MMP)-1) and stromelysin-1 (MMP-3) mRNA levels in cultured human dermal fibroblasts. Journal of Biological Chemistry 273(9): 5279-5287.

Carson KC, Glenn AR, Dilworth MJ. 1994. Specificity of siderophore-mediated transport of iron in rhizobia. Archives of Microbiology 161: 333-339.

Carter RA, Worsley PS, Sawers G, Challis GL, Dilworth MJ, Carson KC, Lawrence JA, Wexler M, Johnston AW, Yeoman KH. 2002. The vbs genes that direct synthesis of the siderophore vicibactin in *Rhizobium leguminosarum*: Their expression in other genera requires ECF sigma factor RpoI. Molecular Microbiology 44(5): 1153-1166.

Chang C, Mooser A, Pluchthun A, Wlodawer A. 2001. Crystal structure of the dimeric C-terminal domain of TonB reveals a novel fold. Journal of Biological Chemistry 276(29): 27535-27540.

Chimento DP, Kadner RJ, Wiener MC. 2005. Comparative structural analysis of TonB-dependent outer membrane transporters: Implications for the transport cycle. Proteins 59(2): 240-251.

Ghosh J, Postle K. 2005. Disulphide trapping of an *in vivo* energy-dependent conformation of *eschericia coli* TonB protein. Molecular Microbiology 55(1): 276-288.

Guerinot ML. 1994. Microbial iron transport. Annual Review Microbiology 48: 743-772.

Gumbart J, Weiner MC, Tajkhorshid E. 2007. Mechanics of force propagation in TonB-dependent outer membrane transport. Journal of Biophysics 93(2): 496-504.

Hantke K. 2001. Iron and metal regulation in bacteria. Current Opinions in Microbiology 4: 172-177.

Higgs PI, Larsen RA, Postle K. 2002. Quantification of the known components of the *Escherichia coli* TonB energy transduction system: TonB, ExbB, ExbD, and FepA. Molecular Microbiology 44(1): 271-281.

Hoang TT, Karkhoff-Schweizer RR, Kutchma AJ, Schweizer HP. 1998. A broad-host-range *flp-FRT* recombination system for site-specific excision of chromosomally-located DNA sequences: Application for isolation of unmarked *Pseudomonas aeruginosa* mutants. Gene 212(1): 77-86.

Horton RM, Cai ZL, Ho SN, Pease LR. 1990. Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. Biotechniques 8*(5)*: 528-535.

Jadhav RS, Desai AJ. 1996. Effect of nutritional and environmental conditions on siderophore production by *Rhizobium* GN1 (peanut isolate). Indian Journal of Experimental Biology 34: 436-439.

James KJ, Hancock MA, Moreau V, Molina F, Coulton JW. 2008. TonB induces conformational changes in surface-exposed loops of FhuA, outer membrane receptor of *Escherichia coli*. Protein Science 17(10): 1879-88.

Kalia A, Rattan A, Chopra P. 1999. A method for extraction of high-quality and high-quantity genomic DNA generally applicable to pathogenic bacteria. Analytical Biochemistry 275: 1-5.

Khursigara CM, De Crescenzo G, Pawelek PD, Coulton JW. 2005. Deletion of the proline-rich region of TonB disrupts formation of a 2:1 complex with FhuA, an outer membrane receptor of *Escherichia coli*. Protein Science 14(5): 1266-1273.

Kneen BE, Larue TA. 1983. Congo red absorption by *Rhizobium leguminosarum*. Applied Environmental Microbiology 45(1): 340-342.

Kodding J, Killig F, Polzer P, Howard SP, Diederichs K, Welte W. 2005. Crystal structure of a 92-residue C-terminal fragment of TonB from *Escherichia coli* reveals significant conformational changes compared to structures of smaller TonB fragments. Journal of Biological Chemistry 280(4): 3022-3028.

Larsen RA, Postle K. 2001. Conserved residues Ser(16) and His(20) and their relative positioning are essential for TonB activity, cross-linking of TonB with ExbB, and the ability of TonB to respond to proton motive force. Journal of Biological Chemistry 276(11): 8111-8117.

Larsen RA, Thomas MG, Postle K. 1999. Protonmotive force, ExbB and ligand-bound FepA drive conformational changes in TonB. Molecular Microbiology 31(6): 1809-1824.

Larsen RA, Wood GE, Postle K. 1993. The conserved proline-rich motif is not essential for energy transduction by *Escherichia coli* TonB protein. Molecular Microbiology 10(5): 943-953.

Long SR. 1996. Rhizobium symbiosis: Nod factors in perspective. Plant Cell 8(10): 1885-1898.

Loomis LD, Raymond KN. 1991. Solution equilibria of enterobactin and metal-enterobactin complexes. Inorganic Chemistry 10: 906-911.

Manhart JR, Wong PP. 1979. Nitrate reductase activities of rhizobia and the correlation between nitrate reduction and nitrogen fixation. Canadian Journal of Microbiology 25: 1169-1174.

McAnulla C, Edwards A, Sanchez-Contreras M, Sawers RG, Downie JA. 2007. Quorum-sensing-regulated transcriptional initiation of plasmid transfer and replication genes in *Rhizobium leguminosarum* biovar *viciae*. Microbiology 153: 2074-2082.

Miethke M, Marahiel MA. 2007. Siderophore-based iron acquisition and pathogen control. Microbiology and Molecular Biology Review 71(3): 413-451.

Neilands JB. 1981. Microbial iron compounds. Annual Review of Biochemistry 50: 715-731.

Ochsner UA, Johnson Z, Vasil ML. 2000. Genetics and regulation of two distinct haem-uptake systems, *phu* and *has*, in *Pseudomonas aeruginosa*. Microbiology 146(1): 185-198.

Ollis AA, Kumar A, Postle K. 2012. The ExbD periplasmic domain contains distinct functional regions for two stages in TonB energization. Journal of Bacteriology 194(12): 3069-77.

Peacock RS, Weijle AM, Howard SP, Price FD, Vogel HJ. 2005. The solution structure of the C-terminal domain of TonB and interaction studies with TonB box peptides. Journal of Molecular Biology 345(4): 1185-1197.

Postle K, Good RF. 1983. DNA sequence of the *Escherichia coli tonB* gene. Proceeding of the National Academy of Sciences of the United States of America 80(17): 5325-5329.

Postle K, Larsen RA. 2007. TonB-dependent energy transduction between outer and cytoplasmic membranes. Biometals 20(3-4): 453-465.

Pramanik A, Hauf W, Hoffmann J, Cemescu M, Brutschy B, Braun V. 2011. Oligomeric structure of ExbB and ExbB-ExbD isolated from *Escherichia coli* as revealed by LILBID mass spectrometry. Biochemistry 50(41): 8950-8956.

Prousek J. 2007. Fenton chemistry in biology and medicine. Pure Applied Chemistry 79(12): 2325-2338.

Ramirez-Bahena MH, Garcia-Fraile P, Peix A, Valverde A, Rivas R, Igual JM, Mateos PF, Martinez-Molina E, Velazquez E. 2008. Revision of the taxonomic status of the species *Rhizobium leguminosarum*(frank 1879) frank 1889^{AL}, *Rhizobium phaseoli* Dangeard 1926^{AL} and *Rhizobium trifolii* Dangeard 1926^{AL}. *R. trifolii* is a later synonym of *R. leguminosarum*. Reclassification of the strain *R. leguminosarum* DSM 30132 (=NCIMB 11478) as *Rhizobium pisi* sp. nov. International Journal of Systematic and Evolutionary Microbiology 58(11): 2484-2490.

Rodionov DA, Gelfand MS, Todd JD, Curson AR, Johnston AW. 2006. Computational reconstruction of iron- and manganese-responsive transcriptional netweorks in alpha-proteobacteria. PLoS Computational Biology 2(12):e163.

Rudnick RL, Gao S. 2003. Composition of the continental crust. Elsevier 3: 1-64.

Shojima S, Nishizawa NK, Fushiya S, Nozoe S, Irifune T, Mori S. 1990. Biosynthesis of phytosiderophores: In vitro biosynthesis of 2'deoxymugineic acid from I-methionine and nicotianamine. Plant Physiology 93(4): 1497-1503.

Shultis DD, Purdy MD, Banchs CN, Wiener MC. 2006. Outer membrane active transport: Structure of BtuB:TonB complex. Science 312(5778): 1396-1399.

Storey E. 2005. Characterization of a Dihydroxamate-Type Siderophore Produced by *Rhizobium leguminosarum* IARI 917: Is it Schizokinen? M. S. Thesis. East Tennessee State University, Johnson City, TN.

Swayne C, Postle K. 2011. Taking the *Escherichia coli* TonB transmembrane domain "offline"? Nonprotonatable asn substitutes fully for TonB His20. Journal of Bacteriology 193(15): 3693-3701.

Todd JD, Sawers G, Rodionov DA, Johnston AW. 2006. The *Rhizobium leguminosarum* regulator IrrA affects the transcription of a wide range of genes in response to Fe availability. Molecular Genetics and Genomics 275(6): 564-577.

Todd JD, Wexler M, Sawers G, Yeoman KH, Poole PS, Johnston AW. 2002. RirA, an iron-responsive regulator in the symbiotic strain *Rhizobium leguminosarum*. Microbiology 148(Pt 12): 4059-4071.

Vakharia-Rao H, Kastead KA, Savenkova MI, Bulathsinghala CM, Postle K. 2007. Deletion and substitution analysis of the *Escherichia coli* TonB Q160 region. Journal of Bacteriology 189(13): 4662-4670.

Walker SA, Downie JA. 2000. Entry of *Rhizobium leguminosarum* bv. *viciae* into root hairs requires minimal nod factor specificity, but subsequent infection thread growth requires *nodO* or *nodE*. Molecular Plant Microbe Interaction 13(7): 754-762.

Weber K, Achenbach LA, Coates JD. 2006. Microbes pumping iron: Anaerobic and microbial iron oxidation and reduction. Nature Reviews Microbiology 4: 752-764.

Wexler M, Yeoman KH, Stevens JB, de Luca NG, Sawers G, Johnston AW. 2001. The *Rhizobium leguminosarum* tonB gene is required for the uptake of siderophore and haem as sources of iron. Molecular Microbiology 41(4): 801-816.

Wright W, Little J, Liu F, Chakraborty RJ. 2013. Isolation and structural identification of the trihydroxamate siderophore vicibactin and its degradative products from *Rhizobium leguminosarum* ATCC 14479 bv. trifolii. Biometals 26(2): 271-83.

Yang X, Hou B, Zong C, Hong G. 2012. Identification of a NodD repressible gene adjacent to *nodM* in *Rhizobium leguminosarum* biovar *viciae*. Acta Biochimica Et Biophysica Sinica 44(4): 323-329.

Yeoman KH, Curson ARJ, Todd JD, Sawers G, Johnston AWB. 2004. Evidence that the rhizobium regulatory protein RirA binds to cis-acting iron-responsive operators (IROs) at promoters of some Fe-regulated genes. Microbiology 150: 4065-4074.

Zhai YF, Heijne W, Saier MHJ. 2003. Molecular modeling of the bacterial outer membrane receptor energizer ExbBD/TonB, based on homology with the flagellar moter, MotAB. Biochemica Et Biophysica Acta- Biomembranes 1614(2): 201-210.

APPENDIX A

MEDIA AND BUFFERS

Modified Manhart and Wong Media

Basal Media	g%
Dextrose	6.0
Glutamate	1.5
KH_2PO_4	1.0
K_2HPO_4	0.764
$MgSO_4$	0.18
$CaSO_4 * 2H_2O$	0.13

Vitamin Solution (1000x concentrated)	mg%
$Na_2EDTA*2H_2O$	550
$Na_2MoO_4*2H_2O$	250
H_3BO_3	145
$ZnSO_4*7H_2O$	108
Calcium Panthenate	50
Inositol	50
Thiamine HCl	40
Biotin	12
$CoCl_2*6H_2O$	10
Riboflavin	10
p-aminobenzoic Acid	10
Nicotinic Acid	10
Pyridoxine HCl	10
Vitamin B12	10
$CuSO_4*5H_2O$	4.37
$MnCl_2*4H_2O$	4.3

SDS-PAGE Solutions and Buffers

Running Gel Buffer pH = 8.8
Tris 36.6 g
ddH_2O to 200 mL

Stacking Gel Buffer pH = 6.8
Tris 3.0 g
ddH_2O to 50 mL

2x Loading Buffer	
Stacking Gel Buffer	2.5 mL
10% SDS	4.0 mL
Glycerol	2.0 mL
2-mercaptoethanol	1.0 mL
Bromophenol blue	2.0 mg
ddH_2O	10 mL

Tank Buffer (4x)	
Tris	12 g
Glycine	57.6 g
SDS	4 g
ddH_2O	to 1L

Coomassie Blue Stock	
Coomassie Blue R-250	2.0 g
ddH_2O	200 mL

Coomassie Stain	
Coomassie Blue Stock	12.5 mL
Methanol	50 mL
Glacial acetic acid	10 mL
ddH_2O	100 mL

Destaining Solution	
Methanol	500 mL
Glacial acetic acid	100 mL
ddH_2O	400 mL

10% SDS-PAGE Recipe

	Volume	
Ingredient	Stacking Gel	Running Gel
30% Bis-acrylamide	1.33 mL	6.66 mL
Running Buffer	0	5.0 mL
Stacking Buffer	2.50 mL	0
ddH_2O	6.1 mL	8.0 mL
10% Ammonium Persulfate	50 µL	100 µL
10% SDS	100 µL	200 µL
TEMED	5 µL	10 µL

Western Blot Solutions

Transfer Buffer	
20mM Tris, pH = 8.3	2.4 g
150mM Glycine	11.26 g
ddH_2O	to 1L

Tris-Buffered Saline (TBS 2x)	
20mM Tris, pH = 7.5	9.7 g
500mM NaCl	117 g
ddH_2O	to 2L

Wash Buffer	
Tween-20	0.5 mL
TBS (1x)	1L

Blocking Buffer	
Bovine Serum Albumin (BSA)	6.0 g
TBS (1x)	to 200 mL

APPENDIX B

SEQUENCE DATA

ATCC 14479 F1-R1 compared to RLΔTB F1-R1

```
CLUSTAL O(1.2.0) multiple sequence alignment


RLTB_001        GCTGATGCCGGTTTCGGCGGCGACGAGGGCTGCCTCGGAAATCTTCAGCTGCGCGGCGAT
ATCC_14479      GCTGATGCCGGTTTCGGCGGCGACGAGGGCTGCCTCGGAAATCTTCAGCTGCGCGGCGAT
                ************************************************************

RLTB_001        ATCGCGCTCGCGCATTTTCGGATTCTCGGCGCGAAACGCACGGATTTCGGCTGGCGCCGG
ATCC_14479      ATCGCGCTCGCGCATTTTCGGATTCTCGGCGCGAAACGCACGGATTTCGGCTGGCGCCGG
                ************************************************************

RLTB_001        TCTTGTCTGTTCGGTCATGTCCTACCCTACTTATTGAGAATGAGCTTGCCCTGACGGGTG
ATCC_14479      TCTTGTCTGTTCGGTCATGTCCTACCCTACTTATTGAGAATGAGCTTGCCCTGACGGGTG
                ************************************************************

RLTB_001        ATCTTCAGGCGATAGACCAAGCCGTCGTGTCTAATCATGATCTCGTTCGTGCCGCGGAAG
ATCC_14479      ATCTTCAGGCGATAGACCAAGCCGTCGTGTCTAATCATGATCTCGTTCGTGCCGCGGAAG
                ************************************************************

RLTB_001        AGATCCGCGCTTTCGACGATCCGGTGCTGCGCCGCAGGCTCGCTCTGCAGCGGCACGTGC
ATCC_14479      AGATCCGCGCTTTCGACGATCCGGTGCTGCGCCGCAGGCTCGCTCTGCAGCGGCACGTGC
                ************************************************************

RLTB_001        TTAAAGTTATCTGGCTTTTCAACCATCATTTCAGTTGGCAATTCCGTGAGACCGGTGATC
ATCC_14479      TTAAAGTTATCTGGCTTTTCAACCATCATTTCAGTTGGCAATTCCGTGAGACCGGTGATC
                ************************************************************

RLTB_001        AGGGTCCCCGGTTGATGACAATTATCTTGACTTTCTTACTCATAGTTTTTTAAAGACGCA
ATCC_14479      AGGGTCCCCGGTTGATGACAATTATCTTGACTTTCTTACTCATAGTTTTTTAAAGACGCA
                ************************************************************

RLTB_001        ATAGGAGACTAATACAGTCAAGTTTTTGAGATTACTGATATGAAGCCGCCTGGCGGCGGT
ATCC_14479      ATAGGAGACTAATACAGTCAAGTTTTTGAGATTACTGATATGAAGCCGCCTGGCGGCGGT
                ************************************************************

RLTB_001        TCGCGACAATGGCGGAGAAGATGATGACGGTTAGGTTTGCGGCTGTTGCAGCCGGCATGG
ATCC_14479      TCGCGACAATGGCGGAGAAGATGATGACGGTTAGGTTTGCGGCTGTTGCAGCCGGCATGG
                ************************************************************

RLTB_001        TGATGGCGGTTTCAAGCGTATGCCAGGCGCAGGATGCGGCAACCGCCAAGCTGGATGTCG
ATCC_14479      TGATGGCGGTTTCAAGCGTATGCCAGGCGCAGGATGCGGCAACCGCCAAGCTGGATGTCG
                ************************************************************

RLTB_001        AACTCACGCTTTGGCGCCGTCGCAAAGGGCTGCATGATGACCTTCGTCGCCGAGAACAAT
ATCC_14479      AACTCACGCTTTGGCGCCGTCGCAAAGGGCTGCATGATGACCTTCGTCGCCGAGAACAAT
                ************************************************************

RLTB_001        CTTCAGGCGCCGATCAACAAGATCTCCTTCGAGCTCGCTTTCTTCAACGACAAGAATGCC
ATCC_14479      CTTCAGGCGCCGATCAACAAGATCTCCTTCGAGCTCGCTTTCTTCAACGACAAGAATGCC
                ************************************************************

RLTB_001        GTCGACCGCATCACCGTGCTCGATTTCCGCGATCTGCCGCAGGGCAAGAAGCGCGTGCGC
ATCC_14479      GTCGACCGCATCACCGTGCTCGATTTCCGCGATCTGCCGCAGGGCAAGAAGCGCGTGCGC
                ************************************************************
```

```
RLTB_001        CAGTTCGATATGCCGAATGTGAAGTGCGAGACGGTGACACGCATCATCATCAACGACACG
ATCC_14479      CAGTTCGATATGCCGAATGTGAAGTGCGAGACGGTGACACGCATCATCATCAACGACACG
                ************************************************************

RLTB_001        CCGGTCTGCGACGGCCCGGCAACGGGCGAATGCATGAAGGGTCTCGTCGCCCGTTCACAG
ATCC_14479      CCGGTCTGCGACGGCCCGGCAACGGGCGAATGCATGAAGGGTCTCGTCGCCCGTTCACAG
                ************************************************************

RLTB_001        ATTTCCGTTCCCTTCGAGGGCTAA-----------------------------------
ATCC_14479      ATTTCCGTTCCCTTCGAGGGCTAAGAATACGCCGGAATGAACCCCGGCGGGCATGGTCAG
                ************************

RLTB_001        ------------------------------------------------------------
ATCC_14479      ACTGTCCGAGGCATGCAAATGGCAATTTCAGCGAAGTCCAGATCGAGACAGGTGCTCATC


RLTB_001        ------------------------------------------------------------
ATCC_14479      GGGGAGCCGGACGCTGACGGCAGTCTGAACGACAATAACATGCATCCCGGCCACGAGCTT


RLTB_001        ------------------------------------------------------------
ATCC_14479      TCCGACTTGCGTAATGTGCAGCGGCAGCCGGCTGGCGAGGCGGTGGTCCACTATGCGCGT


RLTB_001        ------------------------------------------------------------
ATCC_14479      TTCGCGCAGATACCCTCCTTTCCCGATCATCCGGAAGCCGAGCCGATAGCCTCCGTTCCT


RLTB_001        ------------------------------------------------------------
ATCC_14479      GCGCCGCCGATCGATGCGGCGGTGGAAAAGCAGGAAGACGAGAGGAAGCCCATGCGGCGA


RLTB_001        ------------------------------------------------------------
ATCC_14479      CGGATGGCGCTGACTTGTATTGGTTCGTTCTTTTTTCATGCAGGTCTGGTGGTTGCTCTT


RLTB_001        ------------------------------------------------------------
ATCC_14479      AGCGTGTTCATGCCGACGCCGCCGAACGAAGCGATTGAAGATGCTGGTGAGGCTGTCAGC


RLTB_001        ------------------------------------------------------------
ATCC_14479      GTCGTCGTGTATGGCGACTCTGACGTTGACCAGACATCGGCCGGCGATCCAGAGTTAGAG


RLTB_001        ------------------------------------------------------------
ATCC_14479      CGCCAGCCCGAGCAAGTTGCCTCGGAGGAGGTTGAGCCGGATACAGTTCAATCGGAGGAG


RLTB_001        ------------------------------------------------------------
ATCC_14479      GCCACCGAACTAGCTGCCACGACCGTGCCGCCGGAGCAGGCCCAGCCGGTCGAAACTGAA


RLTB_001        ------------------------------------------------------------
ATCC_14479      ACGATGGAGACCGTTCAGTCGGTGCAGGAAGTTACGCGGGTTTCCCCTGAGGCCGTTGTT


RLTB_001        ------------------------------------------------------------
ATCC_14479      GCCGCCGAGCCAGAGGTGCTTGTCTCCGAGAGCCCTGCAGAATCTTTCGTCGCTCAGCCG


RLTB_001        ------------------------------------------------------------
ATCC_14479      ATGGCAACGGCTGTGCCCGAACAGCCCATGCCTGATATCGCGCAGGCAACAGTACCGGAA
```

```
RLTB_001        ------------------------------------------------------------
ATCC_14479      GAAGTAGTGCCAACGGCTGTACAGCCCACCGCAGTTCCGCCGGAAGAAGTAAAGCCTGTC

RLTB_001        ------------------------------------------------------------
ATCC_14479      GAGACTGTCGAAGTGTCGCCGGAACCTGAAGACAAACCGAAGCCGAAGGCTAAAAAACCG

RLTB_001        ------------------------------------------------------------
ATCC_14479      AAACCGGTTGAAAAGAAGCAGTCTCAGAAGAGGGCCAAGCCTGCTGGTGGGAAAGAGGGC

RLTB_001        ------------------------------------------------------------
ATCC_14479      TCAGATCGTGAGGACTCGACGAGGGGCATGGTGAACGGCCAGGACGGGCCGCAAACTGAC

RLTB_001        ------------------------------------------------------------
ATCC_14479      GGCACGTCAACCACGACAGGCGGAACTGACGGTATGGGAAGTGCTGCCGTAGCCAACTAC

RLTB_001        ------------------------------------------------------------
ATCC_14479      CCTGGTAAGATACAGAAGCGGATCCGTCGTGCTGTTCGCGTGCCTGATGAATACAAAAAC

RLTB_001        ------------------------------------------------------------
ATCC_14479      AAGAGCGGCGGTATGACCGTAAGGATTCAATTGACGATTAACGGAACAGGCAGGGTAGCG

RLTB_001        ------------------------------------------------------------
ATCC_14479      TCGGTCTCTGTTGCTCGCAGTTCGGGTATTGCAGAATTGGATAAAGCTGTTCTGGATGGT

RLTB_001        ------------------------------------------------------------G
ATCC_14479      GTGCGGCGCGCGGCGCCGTTTCCGCCGCTGCCCTCGGAGTGGGGTAAGCCCAGCTGGACT

RLTB_001        AATACGCCGGAAGTTCAGGTCACTAGATAATATGATAAAAATAATCAACTTTATACTTTA
ATCC_14479      TTCGCCCAGGAAGTTCAGGTCACTAGATAATATGATAAAAATAATCAACTTTATACTTTA
                :: .* *.**************************************************

RLTB_001        CTCTAAAAATCAAGTTTAATAAGGTCCGCTCGCACACGCAGGAATCGTGTGACGATCAAC
ATCC_14479      CTCTAAAAATCAAGTTTAATAAGGTCCGCTCGCACACGCAGGAATCGTGTGACGATCAAC
                ************************************************************

RLTB_001        GAGCGAACGGGTGGTATATGGCTCGCAAGGGCAAGAAGGGTTCGAACCGGGAGGTCCGCG
ATCC_14479      GAGCGAACGGGTGGTATATGGCTCGCAAGGGCAAGAAGGGTTCGAACCGGGAGGTCCGCG
                ************************************************************

RLTB_001        ACAGCGCGGGCGAGGACGCCGGTCAGGCATCCCTTGGGCGATCCAAACTGGATGGCCGCA
ATCC_14479      ACAGCGCGGGCGAGGACGCCGGTCAGGCATCCCTTGGGCGATCCAAACTGGATGGCCGCA
                ************************************************************

RLTB_001        GTTTCCTCTATGTCGGCGGCCGTGACTGCCAGGTGGCGCATCTTCGTCAGATATGCAGCA
ATCC_14479      GTTTCCTCTATGTCGGCGGCCGTGACTGCCAGGTGGCGCATCTTCGTCAGATATGCAGCA
                ************************************************************

RLTB_001        ATTTCGGTGCAGAGCTCATTCATCACGACGGCGGCTTACGCGAAGCGGTTTCGCGCATCG
ATCC_14479      ATTTCGGTGCAGAGCTCATTCATCACGACGGCGGCTTACGCGAAGCGGTTTCGCGCATCG
                ************************************************************

RLTB_001        ATACCGTCCTTCCCTCGGTCGACTGCGTGTTCTGCCCGATCGACTGTATCAGCCACGATG
```

```
ATCC_14479      ATACCGTCCTTCCCTCGGTCGACTGCGTGTTCTGCCCGATCGACTGTATCAGCCACGATG
                ************************************************************

RLTB_001        CCTGCCTGCGCGTGAAGACCGGCTGCAAGAAGTTCGGCAAGGCCTTCATCCCGCTTCGCA
ATCC_14479      CCTGCCTGCGCGTGAAGACCGGCTGCAAGAAGTTCGGCAAGGCCTTCATCCCGCTTCGCA
                ************************************************************

RLTB_001        ACGGCAGCAAGTCCAGCCTGGAGCGTGCGCTGCAGACGATGAACGAACGAGACAATTCCC
ATCC_14479      ACGGCAGCAAGTCCAGCCTGGAGCGTGCGCTGCAGACGATGAACGAACGAGACAATTCCC
                ************************************************************

RLTB_001        GATGAACGACCAGCGCCCCGACGGCTTCACCATGATCGGCCTGCACAAGCTTGCCGCGCA
ATCC_14479      GATGAACGACCAGCGCCCCGACGGCTTCACCATGATCGGCCTGCACAAGCTTGCCGCGCA
                ************************************************************

RLTB_001        GACAGGCGAGGGCCTCGTGCCCGAACTCTACGAGCTTTTCCAGCAGCATGCCGAACGCCA
ATCC_14479      GACAGGCGAGGGCCTCGTGCCCGAACTCTACGAGCTTTTCCAGCAGCATGCCGAACGCCA
                ************************************************************

RLTB_001        GCGGATCTACCAGAATGTGATGTTCTTTCCGACCTGGGAAGCGCGCATGCCCGGAGAGGC
ATCC_14479      GCGGATCTACCAGAATGTGATGTTCTTTCCGACCTGGGAAGCGCGCATGCCCGGAGAGGC
                ************************************************************

RLTB_001        GGCAACAAGACCGCTTGGCCCGGCGGCGGCAAACGGCGATAATGTTCTCGCCTTTCCGTC
ATCC_14479      GGCAACAAGACCGCTTGGCCCGGCGGCGGCAAACGGCGATAATGTTCTCGCCTTTCCGTC
                ************************************************************

RLTB_001        GCGTGCGGCAAGGGCGGGCAAGAGGAAGGCGTAAGGAGATTTCATGTATATCGCAATGAA
ATCC_14479      GCGTGCGGCAAGGGCGGGCAAGAGGAAGGCGTAAGGAGATTTCATGTATATCGCAATGAA
                ************************************************************

RLTB_001        CCGCTTCAAGGTTGCAACCGGGAGCGAAGGCGATTTCGAGACGGTCTGGCGCAATCGCGA
ATCC_14479      CCGCTTCAAGGTTGCAACCGGGAGCGAAGGCGATTTCGAGACGGTCTGGCGCAATCGCGA
                ************************************************************

RLTB_001        TTCCAGCCTGCCCGAGGTGCCCGGTTTCGTCGAATTCCGCCTGCTGCGCGGCAAGGTCAA
ATCC_14479      TTCCAGCCTGCCCGAGGTGCCCGGTTTCGTCGAATTCCGCCTGCTGCGCGGCAAGGTCAA
                ************************************************************

RLTB_001        CGACGAGGAAGGCTATACGCTCTATTCCTCGCACACGGTCTGGAAAAGCGAAGCG
ATCC_14479      CGACGAGGAAGGCTATACGCTCTATTCCTCGCACACGGTCTGGAAAAGCGAAGCG
                ********************************************************
```

ATCC 14479 F1-R1 compared to pMRTP1 F1-R1

CLUSTAL O(1.2.0) multiple sequence alignment

```
ATCC_14479      GCTGATGCCGGTTTCGGCGGCGACGAGGGCTGCCTCGGAAATCTTCAGCTGCGCGGCGAT
pMRTP_1         GCTGATGCCGGTTTCGGCGGCGACGAGGGCTGCCTCGGAAATCTTCAGCTGCGCGGCGAT
                ************************************************************

ATCC_14479      ATCGCGCTCGCGCATTTTCGGATTCTCGGCGCGAAACGCACGGATTTCGGCTGGCGCCGG
pMRTP_1         ATCGCGCTCGCGCATTTTCGGATTCTCGGCGCGAAACGCACGGATTTCGGCTGGCGCCGG
                ************************************************************

ATCC_14479      TCTTGTCTGTTCGGTCATGTCCTACCCTACTTATTGAGAATGAGCTTGCCCTGACGGGTG
pMRTP_1         TCTTGTCTGTTCGGTCATGTCCTACCCTACTTATTGAGAATGAGCTTGCCCTGACGGGTG
```

```
                ************************************************************

ATCC_14479      ATCTTCAGGCGATAGACCAAGCCGTCGTGTCTAATCATGATCTCGTTCGTGCCGCGGAAG
pMRTP_1         ATCTTCAGGCGATAGACCAAGCCGTCGTGTCTAATCATGATCTCGTTCGTGCCGCGGAAG
                ************************************************************

ATCC_14479      AGATCCGCGCTTTCGACGATCCGGTGCTGCGCCGCAGGCTCGCTCTGCAGCGGCACGTGC
pMRTP_1         AGATCCGCGCTTTCGACGATCCGGTGCTGCGCCGCAGGCTCGCTCTGCAGCGGCACGTGC
                ************************************************************

ATCC_14479      TTAAAGTTATCTGGCTTTTCAACCATCATTTCAGTTGGCAATTCCGTGAGACCGGTGATC
pMRTP_1         TTAAAGTTATCTGGCTTTTCAACCATCATTTCAGTTGGCAATTCCGTGAGACCGGTGATC
                ************************************************************

ATCC_14479      AGGGTCCCCGGTTGATGACAATTATCTTGACTTTCTTACTCATAGTTTTTTAAAGACGCA
pMRTP_1         AGGGTCCCCGGTTGATGACAATTATCTTGACTTTCTTACTCATAGTTTTTTAAAGACGCA
                ************************************************************

ATCC_14479      ATAGGAGACTAATACAGTCAAGTTTTTGAGATTACTGATATGAAGCCGCCTGGCGGCGGT
pMRTP_1         ATAGGAGACTAATACAGTCAAGTTTTTGAGATTACTGATATGAAGCCGCCTGGCGGGAGG
                ************************************************************ .*

ATCC_14479      TCGCGACAATGGCGGAGAAGATGATGACGGTTAGGTTTGCGGCTGTTGCAGCCGGCATGG
pMRTP_1         G------------------------------------------------------------

ATCC_14479      TGATGGCGGTTTCAAGCGTATGCCAGGCGCAGGATGCGGCAACCGCCAAGCTGGATGTCG
pMRTP_1         ------------------------------------------------------------

ATCC_14479      AACTCACGCTTTGGCGCCGTCGCAAAGGGCTGCATGATGACCTTCGTCGCCGAGAACAAT
pMRTP_1         ------------------------------------------------------------

ATCC_14479      CTTCAGGCGCCGATCAACAAGATCTCCTTCGAGCTCGCTTTCTTCAACGACAAGAATGCC
pMRTP_1         ------------------------------------------------------------

ATCC_14479      GTCGACCGCATCACCGTGCTCGATTTCCGCGATCTGCCGCAGGGCAAGAAGCGCGTGCGC
pMRTP_1         ------------------------------------------------------------

ATCC_14479      CAGTTCGATATGCCGAATGTGAAGTGCGAGACGGTGACACGCATCATCATCAACGACACG
pMRTP_1         ------------------------------------------------------------

ATCC_14479      CCGGTCTGCGACGGCCCGGCAACGGGCGAATGCATGAAGGGTCTCGTCGCCCGTTCACAG
pMRTP_1         ------------------------------------------------------------

ATCC_14479      ATTTCCGTTCCCTTCGAGGGCTAAGAATACGCCGGAATGAACCCCGGCGGGCATGGTCAG
pMRTP_1         --------------------CTAAGAATACGCCGGAATGAACCCCGGCGGGCATGGTCAG
                                    ************************************

ATCC_14479      ACTGTCCGAGGCATGCAAATGGCAATTTCAGCGAAGTCCAGATCGAGACAGGTGCTCATC
pMRTP_1         ACTGTCCGAGGCATGCAAATGGCAATTTCAGCGAAGTCCAGATCGAGACAGGTGCTCATC
                ************************************************************

ATCC_14479      GGGGAGCCGGACGCTGACGGCAGTCTGAACGACAATAACATGCATCCCGGCCACGAGCTT
pMRTP_1         GGGGAGCCGGACGCTGACGGCAGTCTGAACGACAATAACATGCATCCCGGCCACGAGCTT
                ************************************************************

ATCC_14479      TCCGACTTGCGTAATGTGCAGCGGCAGCCGGCTGGCGAGGCGGTGGTCCACTATGCGCGT
```

```
pMRTP_1         TCCGACTTGCGTAATGTGCAGCGGCAGCCGGCTGGCGAGGCGGTGGTCCACTATGCGCGT
                ************************************************************

ATCC_14479      TTCGCGCAGATACCCTCCTTTCCCGATCATCCGGAAGCCGAGCCGATAGCCTCCGTTCCT
pMRTP_1         TTCGCGCAGATACCCTCCTTTCCCGATCATCCGGAAGCCGAGCCGATAGCCTCCGTTCCT
                ************************************************************

ATCC_14479      GCGCCGCCGATCGATGCGGCGGTGGAAAAGCAGGAAGACGAGAGGAAGCCCATGCGGCGA
pMRTP_1         GCGCCGCCGATCGATGCGGCGGTGGAAAAGCAGGAAGACGAGAGGAAGCCCATGCGGCGA
                ************************************************************

ATCC_14479      CGGATGGCGCTGACTTGTATTGGTTCGTTCTTTTTTCATGCAGGTCTGGTGGTTGCTCTT
pMRTP_1         CGGATGGCGCTGACTTGTATTGGTTCGTTCTTTTTTCATGCAGGTCTGGTGGTTGCTCTT
                ************************************************************

ATCC_14479      AGCGTGTTCATGCCGACGCCGCCGAACGAAGCGATTGAAGATGCTGGTGAGGCTGTCAGC
pMRTP_1         AGCGTGTTCATGCCGACGCCGCCGAACGAAGCGATTGAAGATGCTGGTGAGGCTGTCAGC
                ************************************************************

ATCC_14479      GTCGTCGTGTATGGCGACTCTGACGTTGACCAGACATCGGCCGGCGATCCAGAGTTAGAG
pMRTP_1         GTCGTCGTGTATGGCGACTCTGACGTTGACCAGACATCGGCCGGCGATCCAGAGTTAGAG
                ************************************************************

ATCC_14479      CGCCAGCCCGAGCAAGTTGCCTCGGAGGAGGTTGAGCCGGATACAGTTCAATCGGAGGAG
pMRTP_1         CGCCAGCCCGAGCAAGTTGCCTCGGAGGAGGTTGAGCCGGATACAGTTCAATCGGAGGAG
                ************************************************************

ATCC_14479      GCCACCGAACTAGCTGCCACGACCGTGCCGCCGGAGCAGGCCCAGCCGGTCGAAACTGAA
pMRTP_1         GCCACCGAACTAGCTGCCACGACCGTGCCGCCGGAGCAGGCCCAGCCGGTCGAAACTGAA
                ************************************************************

ATCC_14479      ACGATGGAGACCGTTCAGTCGGTGCAGGAAGTTACGCGGGTTTCCCCTGAGGCCGTTGTT
pMRTP_1         ACGATGGAGACCGTTCAGTCGGTGCAGGAAGTTACGCGGGTTTCCCCTGAGGCCGTTGTT
                ************************************************************

ATCC_14479      GCCGCCGAGCCAGAGGTGCTTGTCTCCGAGAGCCCTGCAGAATCTTTCGTCGCTCAGCCG
pMRTP_1         GCCGCCGAGCCAGAGGTGCTTGTCTCCGAGAGCCCTGCAGAATCTTTCGTCGCTCAGCCG
                ************************************************************

ATCC_14479      ATGGCAACGGCTGTGCCCGAACAGCCCATGCCTGATATCGCGCAGGCAACAGTACCGGAA
pMRTP_1         ATGGCAACGGCTGTGCCCGAACAGCCCATGCCTGATATCGCGCAGGCAACAGTACCGGAA
                ************************************************************

ATCC_14479      GAAGTAGTGCCAACGGCTGTACAGCCCACCGCAGTTCCGCCGGAAGAAGTAAAGCCTGTC
pMRTP_1         GAAGTAGTGCCAACGGCTGTACAGCCCACCGCAGTTCCGCCGGAAGAAGTAAAGCCTGTC
                ************************************************************

ATCC_14479      GAGACTGTCGAAGTGTCGCCGGAACCTGAAGACAAACCGAAGCCGAAGGCTAAAAAACCG
pMRTP_1         GAGACTGTCGAAGTGTCGCCGGAACCTGAAGACAAACCGAAGCCGAAGGCTAAAAAACCG
                ************************************************************

ATCC_14479      AAACCGGTTGAAAAGAAGCAGTCTCAGAAGAGGGCCAAGCCTGCTGGTGGGAAAGAGGGC
pMRTP_1         AAACCGGTTGAAAAGAAGCAGTCTCAGAAGAGGGCCAAGCCTGCTGGTGGGAAAGAGGGC
                ************************************************************

ATCC_14479      TCAGATCGTGAGGACTCGACGAGGGGCATGGTGAACGGCCAGGACGGGCCGCAAACTGAC
pMRTP_1         TCAGATCGTGAGGACTCGACGAGGGGCATGGTGAACGGCCAGGACGGGCCGCAAACTGAC
                ************************************************************

ATCC_14479      GGCACGTCAACCACGACAGGCGGAACTGACGGTATGGGAAGTGCTGCCGTAGCCAACTAC
pMRTP_1         GGCACGTCAACCACGACAGGCGGAACTGACGGTATGGGAAGTGCTGCCGTAGCCAACTAC
                ************************************************************
```

```
ATCC_14479      CCTGGTAAGATACAGAAGCGGATCCGTCGTGCTGTTCGCGTGCCTGATGAATACAAAAAC
pMRTP_1         CCTGGTAAGATACAGAAGCGGATCCGTCGTGCTGTTCGCGTGCCTGATGAATACAAAAAC
                ************************************************************

ATCC_14479      AAGAGCGGCGGTATGACCGTAAGGATTCAATTGACGATTAACGGAACAGGCAGGGTAGCG
pMRTP_1         AAGAGCGGCGGTATGACCGTAAGGATTCAATTGACGATTAACGGAACAGGCAGGGTAGCG
                ************************************************************

ATCC_14479      TCGGTCTCTGTTGCTCGCAGTTCGGGTATTGCAGAATTGGATAAAGCTGTTCTGGATGGT
pMRTP_1         TCGGTCTCTGTTGCTCGCAGTTCGGGTATTGCAGAATTGGATAAAGCTGTTCTGGATGGT
                ************************************************************

ATCC_14479      GTGCGGCGCGCGGCGCCGTTTCCGCCGCTGCCCTCGGAGTGGGGTAAGCCCAGCTGGACT
pMRTP_1         GTGCGGCGCGCGGCGCCGTTTCCGCCGCTGCCCTCGGAGTGGGGTAAGCCCAGCTGGACT
                ************************************************************

ATCC_14479      TTCGCCCAGGAAGTTCAGGTCACTAGATAATATGATAAAAATAATCAACTTTATACTTTA
pMRTP_1         TTCGCCCAGGAAGTTCAGGTCACTAGATAATATGATAAAAATAATCAACTTTATACTTTA
                ************************************************************

ATCC_14479      CTCTAAAAATCAAGTTTAATAAGGTCCGCTCGCACACGCAGGAATCGTGTGACGATCAAC
pMRTP_1         CTCTAAAAATCAAGTTTAATAAGGTCCGCTCGCACACGCAGGAATCGTGTGACGATCAAC
                ************************************************************

ATCC_14479      GAGCGAACGGGTGGTATATGGCTCGCAAGGGCAAGAAGGGTTCGAACCGGGAGGTCCGCG
pMRTP_1         GAGCGAACGGGTGGTATATGGCTCGCAAGGGCAAGAAGGGTTCGAACCGGGAGGTCCGCG
                ************************************************************

ATCC_14479      ACAGCGCGGGCGAGGACGCCGGTCAGGCATCCCTTGGGCGATCCAAACTGGATGGCCGCA
pMRTP_1         ACAGCGCGGGCGAGGACGCCGGTCAGGCATCCCTTGGGCGATCCAAACTGGATGGCCGCA
                ************************************************************

ATCC_14479      GTTTCCTCTATGTCGGCGGCCGTGACTGCCAGGTGGCGCATCTTCGTCAGATATGCAGCA
pMRTP_1         GTTTCCTCTATGTCGGCGGCCGTGACTGCCAGGTGGCGCATCTTCGTCAGATATGCAGCA
                ************************************************************

ATCC_14479      ATTTCGGTGCAGAGCTCATTCATCACGACGGCGGCTTACGCGAAGCGGTTTCGCGCATCG
pMRTP_1         ATTTCGGTGCAGAGCTCATTCATCACGACGGCGGCTTACGCGAAGCGGTTTCGCGCATCG
                ************************************************************

ATCC_14479      ATACCGTCCTTCCCTCGGTCGACTGCGTGTTCTGCCCGATCGACTGTATCAGCCACGATG
pMRTP_1         ATACCGTCCTTCCCTCGGTCGACTGCGTGTTCTGCCCGATCGACTGTATCAGCCACGATG
                ************************************************************

ATCC_14479      CCTGCCTGCGCGTGAAGACCGGCTGCAAGAAGTTCGGCAAGGCCTTCATCCCGCTTCGCA
pMRTP_1         CCTGCCTGCGCGTGAAGACCGGCTGCAAGAAGTTCGGCAAGGCCTTCATCCCGCTTCGCA
                ************************************************************

ATCC_14479      ACGGCAGCAAGTCCAGCCTGGAGCGTGCGCTGCAGACGATGAACGAACGAGACAATTCCC
pMRTP_1         ACGGCAGCAAGTCCAGCCTGGAGCGTGCGCTGCAGACGATGAACGAACGAGACAATTCCC
                ************************************************************

ATCC_14479      GATGAACGACCAGCGCCCCGACGGCTTCACCATGATCGGCCTGCACAAGCTTGCCGCGCA
pMRTP_1         GATGAACGACCAGCGCCCCGACGGCTTCACCATGATCGGCCTGCACAAGCTTGCCGCGCA
                ************************************************************

ATCC_14479      GACAGGCGAGGGCCTCGTGCCCGAACTCTACGAGCTTTTCCAGCAGCATGCCGAACGCCA
pMRTP_1         GACAGGCGAGGGCCTCGTGCCCGAACTCTACGAGCTTTTCCAGCAGCATGCCGAACGCCA
                ************************************************************

ATCC_14479      GCGGATCTACCAGAATGTGATGTTCTTTCCGACCTGGGAAGCGCGCATGCCCGGAGAGGC
pMRTP_1         GCGGATCTACCAGAATGTGATGTTCTTTCCGACCTGGGAAGCGCGCATGCCCGGAGAGGC
                ************************************************************
```

```
ATCC_14479      GGCAACAAGACCGCTTGGCCCGGCGGCGGCAAACGGCGATAATGTTCTCGCCTTTCCGTC
pMRTP_1         GGCAACAAGACCGCTTGGCCCGGCGGCGGCAAACGGCGATAATGTTCTCGCCTTTCCGTC
                ************************************************************

ATCC_14479      GCGTGCGGCAAGGGCGGGCAAGAGGAAGGCGTAAGGAGATTTCATGTATATCGCAATGAA
pMRTP_1         GCGTGCGGCAAGGGCGGGCAAGAGGAAGGCGTAAGGAGATTTCATGTATATCGCAATGAA
                ************************************************************

ATCC_14479      CCGCTTCAAGGTTGCAACCGGGAGCGAAGGCGATTTCGAGACGGTCTGGCGCAATCGCGA
pMRTP_1         CCGCTTCAAGGTTGCAACCGGGAGCGAAGGCGATTTCGAGACGGTCTGGCGCAATCGCGA
                ************************************************************

ATCC_14479      TTCCAGCCTGCCCGAGGTGCCCGGTTTCGTCGAATTCCGCCTGCTGCGCGGCAAGGTCAA
pMRTP_1         TTCCAGCCTGCCCGAGGTGCCCGGTTTCGTCGAATTCCGCCTGCTGCGCGGCAAGGTCAA
                ************************************************************

ATCC_14479      CGACGAGGAAGGCTATACGCTCTATTCCTCGCACACGGTCTGGAAAAGCGAAGCG
pMRTP_1         CGACGAGGAAGGCTATACGCTCTATTCCTCGCACACGGTCTGGAAAAGCGAAGCG
                ***********************************************************
```

ATCC 14479 F1-R1 compared to pMRTP2 F1-R1

```
CLUSTAL 2.1 multiple sequence alignment

ATCC_14479      GCTGATGCCGGTTTCGGCGGCGACGAGGGCTGCCTCGGAAATCTTCAGCTGCGCGGCGAT
pMRTP_2         GCTGATGCCGGTTTCGGCGGCGACGAGGGCTGCCTCGGAAATCTTCAGCTGCGCGGCGAT
                ************************************************************

ATCC_14479      ATCGCGCTCGCGCATTTTCGGATTCTCGGCGCGAAACGCACGGATTTCGGCTGGCGCCGG
pMRTP_2         ATCGCGCTCGCGCATTTTCGGATTCTCGGCGCGAAACGCACGGATTTCGGCTGGCGCCGG
                ************************************************************

ATCC_14479      TCTTGTCTGTTCGGTCATGTCCTACCCTACTTATTGAGAATGAGCTTGCCCTGACGGGTG
pMRTP_2         TCTTGTCTGTTCGGTCATGTCCTACCCTACTTATTGAGAATGAGCTTGCCCTGACGGGTG
                ************************************************************

ATCC_14479      ATCTTCAGGCGATAGACCAAGCCGTCGTGTCTAATCATGATCTCGTTCGTGCCGCGGAAG
pMRTP_2         ATCTTCAGGCGATAGACCAAGCCGTCGTGTCTAATCATGATCTCGTTCGTGCCGCGGAAG
                ************************************************************

ATCC_14479      AGATCCGCGCTTTCGACGATCCGGTGCTGCGCCGCAGGCTCGCTCTGCAGCGGCACGTGC
pMRTP_2         AGATCCGCGCTTTCGACGATCCGGTGCTGCGCCGCAGGCTCGCTCTGCAGCGGCACGTGC
                ************************************************************

ATCC_14479      TTAAAGTTATCTGGCTTTTCAACCATCATTTCAGTTGGCAATTCCGTGAGACCGGTGATC
pMRTP_2         TTAAAGTTATCTGGCTTTTCAACCATCATTTCAGTTGGCAATTCCGTGAGACCGGTGATC
                ************************************************************

ATCC_14479      AGGGTCCCCGGTTGATGACAATTATCTTGACTTTCTTACTCATAGTTTTTTAAAGACGCA
pMRTP_2         AGGGTCCCCGGTTGATGACAATTATCTTGACTTTCTTACTCATAGTTTTTTAAAGACGCA
                ************************************************************

ATCC_14479      ATAGGAGACTAATACAGTCAAGTTTTTGAGATTACTGATATGAAGCCGCCTGGCGGCGGT
pMRTP_2         ATAGGAGACTAATACAGTCAAGTTTTTGAGATTACTGATATGAAGCCGCCTGGCGG----
                ************************************************         

ATCC_14479      TCGCGACAATGGCGGAGAAGATGATGACGGTTAGGTTTGCGGCTGTTGCAGCCGGCATGG
pMRTP_2         ------------------------------------------------------------
```

```
ATCC_14479      TGATGGCGGTTTCAAGCGTATGCCAGGCGCAGGATGCGGCAACCGCCAAGCTGGATGTCG
pMRTP_2         ------------------------------------------------------------

ATCC_14479      AACTCACGCTTTGGCGCCGTCGCAAAGGGCTGCATGATGACCTTCGTCGCCGAGAACAAT
pMRTP_2         ------------------------------------------------------------

ATCC_14479      CTTCAGGCGCCGATCAACAAGATCTCCTTCGAGCTCGCTTTCTTCAACGACAAGAATGCC
pMRTP_2         ------------------------------------------------------------

ATCC_14479      GTCGACCGCATCACCGTGCTCGATTTCCGCGATCTGCCGCAGGGCAAGAAGCGCGTGCGC
pMRTP_2         ------------------------------------------------------------

ATCC_14479      CAGTTCGATATGCCGAATGTGAAGTGCGAGACGGTGACACGCATCATCATCAACGACACG
pMRTP_2         ------------------------------------------------------------

ATCC_14479      CCGGTCTGCGACGGCCCGGCAACGGGCGAATGCATGAAGGGTCTCGTCGCCCGTTCACAG
pMRTP_2         ------------------------------------------------------------

ATCC_14479      ATTTCCGTTCCCTTCGAGGGCTAAGAATACGCCGGAATGAACCCCGGCGGGCATGGTCAG
pMRTP_2         ------------------------------------------------------------

ATCC_14479      ACTGTCCGAGGCATGCAAATGGCAATTTCAGCGAAGTCCAGATCGAGACAGGTGCTCATC
pMRTP_2         ------------------------------------------------------------

ATCC_14479      GGGGAGCCGGACGCTGACGGCAGTCTGAACGACAATAACATGCATCCCGGCCACGAGCTT
pMRTP_2         ------------------------------------------------------------

ATCC_14479      TCCGACTTGCGTAATGTGCAGCGGCAGCCGGCTGGCGAGGCGGTGGTCCACTATGCGCGT
pMRTP_2         ------------------------------------------------------------

ATCC_14479      TTCGCGCAGATACCCTCCTTTCCCGATCATCCGGAAGCCGAGCCGATAGCCTCCGTTCCT
pMRTP_2         ------------------------------------------------------------

ATCC_14479      GCGCCGCCGATCGATGCGGCGGTGGAAAAGCAGGAAGACGAGAGGAAGCCCATGCGGCGA
pMRTP_2         ------------------------------GCAGGAAGACGAGAGGAAGCCCATGCGGCGA
                                               ******************************

ATCC_14479      CGGATGGCGCTGACTTGTATTGGTTCGTTCTTTTTTCATGCAGGTCTGGTGGTTGCTCTT
pMRTP_2         CGGATGGCGCTGACTTGTATTGGTTCGTTCTTTTTTCATGCAGGTCTGGTGGTTGCTCTT
                ************************************************************

ATCC_14479      AGCGTGTTCATGCCGACGCCGCCGAACGAAGCGATTGAAGATGCTGGTGAGGCTGTCAGC
pMRTP_2         AGCGTGTTCATGCCGACGCCGCCGAACGAAGCGATTGAAGATGCTGGTGAGGCTGTCAGC
                ************************************************************

ATCC_14479      GTCGTCGTGTATGGCGACTCTGACGTTGACCAGACATCGGCCGGCGATCCAGAGTTAGAG
pMRTP_2         GTCGTCGTGTATGGCGACTCTGACGTTGACCAGACATCGGCCGGCGATCCAGAGTTAGAG
                ************************************************************

ATCC_14479      CGCCAGCCCGAGCAAGTTGCCTCGGAGGAGGTTGAGCCGGATACAGTTCAATCGGAGGAG
pMRTP_2         CGCCAGCCCGAGCAAGTTGCCTCGGAGGAGGTTGAGCCGGATACAGTTCAATCGGAGGAG
```

```
                ************************************************************

ATCC_14479      GCCACCGAACTAGCTGCCACGACCGTGCCGCCGGAGCAGGCCCAGCCGGTCGAAACTGAA
pMRTP_2         GCCACCGAACTAGCTGCCACGACCGTGCCGCCGGAGCAGGCCCAGCCGGTCGAAACTGAA
                ************************************************************

ATCC_14479      ACGATGGAGACCGTTCAGTCGGTGCAGGAAGTTACGCGGGTTTCCCCTGAGGCCGTTGTT
pMRTP_2         ACGATGGAGACCGTTCAGTCGGTGCAGGAAGTTACGCGGGTTTCCCCTGAGGCCGTTGTT
                ************************************************************

ATCC_14479      GCCGCCGAGCCAGAGGTGCTTGTCTCCGAGAGCCCTGCAGAATCTTTCGTCGCTCAGCCG
pMRTP_2         GCCGCCGAGCCAGAGGTGCTTGTCTCCGAGAGCCCTGCAGAATCTTTCGTCGCTCAGCCG
                ************************************************************

ATCC_14479      ATGGCAACGGCTGTGCCCGAACAGCCCATGCCTGATATCGCGCAGGCAACAGTACCGGAA
pMRTP_2         ATGGCAACGGCTGTGCCCGAACAGCCCATGCCTGATATCGCGCAGGCAACAGTACCGGAA
                ************************************************************

ATCC_14479      GAAGTAGTGCCAACGGCTGTACAGCCCACCGCAGTTCCGCCGGAAGAAGTAAAGCCTGTC
pMRTP_2         GAAGTAGTGCCAACGGCTGTACAGCCCACCGCAGTTCCGCCGGAAGAAGTAAAGCCTGTC
                ************************************************************

ATCC_14479      GAGACTGTCGAAGTGTCGCCGGAACCTGAAGACAAACCGAAGCCGAAGGCTAAAAAACCG
pMRTP_2         GAGACTGTCGAAGTGTCGCCGGAACCTGAAGACAAACCGAAGCCGAAGGCTAAAAAACCG
                ************************************************************

ATCC_14479      AAACCGGTTGAAAAGAAGCAGTCTCAGAAGAGGGCCAAGCCTGCTGGTGGGAAAGAGGGC
pMRTP_2         AAACCGGTTGAAAAGAAGCAGTCTCAGAAGAGGGCCAAGCCTGCTGGTGGGAAAGAGGGC
                ************************************************************

ATCC_14479      TCAGATCGTGAGGACTCGACGAGGGGCATGGTGAACGGCCAGGACGGGCCGCAAACTGAC
pMRTP_2         TCAGATCGTGAGGACTCGACGAGGGGCATGGTGAACGGCCAGGACGGGCCGCAAACTGAC
                ************************************************************

ATCC_14479      GGCACGTCAACCACGACAGGCGGAACTGACGGTATGGGAAGTGCTGCCGTAGCCAACTAC
pMRTP_2         GGCACGTCAACCACGACAGGCGGAACTGACGGTATGGGAAGTGCTGCCGTAGCCAACTAC
                ************************************************************

ATCC_14479      CCTGGTAAGATACAGAAGCGGATCCGTCGTGCTGTTCGCGTGCCTGATGAATACAAAAAC
pMRTP_2         CCTGGTAAGATACAGAAGCGGATCCGTCGTGCTGTTCGCGTGCCTGATGAATACAAAAAC
                ************************************************************

ATCC_14479      AAGAGCGGCGGTATGACCGTAAGGATTCAATTGACGATTAACGGAACAGGCAGGGTAGCG
pMRTP_2         AAGAGCGGCGGTATGACCGTAAGGATTCAATTGACGATTAACGGAACAGGCAGGGTAGCG
                ************************************************************

ATCC_14479      TCGGTCTCTGTTGCTCGCAGTTCGGGTATTGCAGAATTGGATAAAGCTGTTCTGGATGGT
pMRTP_2         TCGGTCTCTGTTGCTCGCAGTTCGGGTATTGCAGAATTGGATAAAGCTGTTCTGGATGGT
                ************************************************************

ATCC_14479      GTGCGGCGCGCGGCGCCGTTTCCGCCGCTGCCCTCGGAGTGGGGTAAGCCCAGCTGGACT
pMRTP_2         GTGCGGCGCGCGGCGCCGTTTCCGCCGCTGCCCTCGGAGTGGGGTAAGCCCAGCTGGACT
                ************************************************************

ATCC_14479      TTCGCCCAGGAAGTTCAGGTCACTAGATAATATGATAAAAATAATCAACTTTATACTTTA
pMRTP_2         TTCGCCCAGGAAGTTCAGGTCACTAGATAATATGATAAAAATAATCAACTTTATACTTTA
                ************************************************************

ATCC_14479      CTCTAAAAATCAAGTTTAATAAGGTCCGCTCGCACACGCAGGAATCGTGTGACGATCAAC
pMRTP_2         CTCTAAAAATCAAGTTTAATAAGGTCCGCTCGCACACGCAGGAATCGTGTGACGATCAAC
                ************************************************************

ATCC_14479      GAGCGAACGGGTGGTATATGGCTCGCAAGGGCAAGAAGGGTTCGAACCGGGAGGTCCGCG
```

```
pMRTP_2         GAGCGAACGGGTGGTATATGGCTCGCAAGGGCAAGAAGGGTTCGAACCGGGAGGTCCGCG
                ************************************************************

ATCC_14479      ACAGCGCGGGCGAGGACGCCGGTCAGGCATCCCTTGGGCGATCCAAACTGGATGGCCGCA
pMRTP_2         ACAGCGCGGGCGAGGACGCCGGTCAGGCATCCCTTGGGCGATCCAAACTGGATGGCCGCA
                ************************************************************

ATCC_14479      GTTTCCTCTATGTCGGCGGCCGTGACTGCCAGGTGGCGCATCTTCGTCAGATATGCAGCA
pMRTP_2         GTTTCCTCTATGTCGGCGGCCGTGACTGCCAGGTGGCGCATCTTCGTCAGATATGCAGCA
                ************************************************************

ATCC_14479      ATTTCGGTGCAGAGCTCATTCATCACGACGGCGGCTTACGCGAAGCGGTTTCGCGCATCG
pMRTP_2         ATTTCGGTGCAGAGCTCATTCATCACGACGGCGGCTTACGCGAAGCGGTTTCGCGCATCG
                ************************************************************

ATCC_14479      ATACCGTCCTTCCCTCGGTCGACTGCGTGTTCTGCCCGATCGACTGTATCAGCCACGATG
pMRTP_2         ATACCGTCCTTCCCTCGGTCGACTGCGTGTTCTGCCCGATCGACTGTATCAGCCACGATG
                ************************************************************

ATCC_14479      CCTGCCTGCGCGTGAAGACCGGCTGCAAGAAGTTCGGCAAGGCCTTCATCCCGCTTCGCA
pMRTP_2         CCTGCCTGCGCGTGAAGACCGGCTGCAAGAAGTTCGGCAAGGCCTTCATCCCGCTTCGCA
                ************************************************************

ATCC_14479      ACGGCAGCAAGTCCAGCCTGGAGCGTGCGCTGCAGACGATGAACGAACGAGACAATTCCC
pMRTP_2         ACGGCAGCAAGTCCAGCCTGGAGCGTGCGCTGCAGACGATGAACGAACGAGACAATTCCC
                ************************************************************

ATCC_14479      GATGAACGACCAGCGCCCCGACGGCTTCACCATGATCGGCCTGCACAAGCTTGCCGCGCA
pMRTP_2         GATGAACGACCAGCGCCCCGACGGCTTCACCATGATCGGCCTGCACAAGCTTGCCGCGCA
                ************************************************************

ATCC_14479      GACAGGCGAGGGCCTCGTGCCCGAACTCTACGAGCTTTTCCAGCAGCATGCCGAACGCCA
pMRTP_2         GACAGGCGAGGGCCTCGTGCCCGAACTCTACGAGCTTTTCCAGCAGCATGCCGAACGCCA
                ************************************************************

ATCC_14479      GCGGATCTACCAGAATGTGATGTTCTTTCCGACCTGGGAAGCGCGCATGCCCGGAGAGGC
pMRTP_2         GCGGATCTACCAGAATGTGATGTTCTTTCCGACCTGGGAAGCGCGCATGCCCGGAGAGGC
                ************************************************************

ATCC_14479      GGCAACAAGACCGCTTGGCCCGGCGGCGGCAAACGGCGATAATGTTCTCGCCTTTCCGTC
pMRTP_2         GGCAACAAGACCGCTTGGCCCGGCGGCGGCAAACGGCGATAATGTTCTCGCCTTTCCGTC
                ************************************************************

ATCC_14479      GCGTGCGGCAAGGGCGGGCAAGAGGAAGGCGTAAGGAGATTTCATGTATATCGCAATGAA
pMRTP_2         GCGTGCGGCAAGGGCGGGCAAGAGGAAGGCGTAAGGAGATTTCATGTATATCGCAATGAA
                ************************************************************

ATCC_14479      CCGCTTCAAGGTTGCAACCGGGAGCGAAGGCGATTTCGAGACGGTCTGGCGCAATCGCGA
pMRTP_2         CCGCTTCAAGGTTGCAACCGGGAGCGAAGGCGATTTCGAGACGGTCTGGCGCAATCGCGA
                ************************************************************

ATCC_14479      TTCCAGCCTGCCCGAGGTGCCCGGTTTCGTCGAATTCCGCCTGCTGCGCGGCAAGGTCAA
pMRTP_2         TTCCAGCCTGCCCGAGGTGCCCGGTTTCGTCGAATTCCGCCTGCTGCGCGGCAAGGTCAA
                ************************************************************

ATCC_14479      CGACGAGGAAGGCTATACGCTCTATTCCTCGCACACGGTCTGGAAAAGCGAAGCG
pMRTP_2         CGACGAGGAAGGCTATACGCTCTATTCCTCGCACACGGTCTGGAAAAGCGAAGCG
                ********************************************************
```

VITA

BRIAN D. HILL

Education:	Gate City High School, Gate City, Virginia B.S., Biology, James Madison University, Harrisonburg, Virginia, 2005 M.S., Biology, East Tennessee State University, Johnson City, Tennessee, 2014
Professional Experience:	Maptech, Inc., Blacksburg, Virginia, 2006-2007 Scott County Department of Recreation, Gate City, Virginia, 2007-2008 Eastman Chemical Company, Kingsport, Tennessee 2008 - 2011 Graduate Assistant, East Tennessee State University, Johnson City, Tennessee, 2011-2014
Honors and Awards:	2012 Appalachian Student Research Forum, 1st Place Poster Presentation, East Tennessee State University