Understanding the Role of a Hemerythrin-Like Protein in Mycobacterium Tumerculosis

Caitlyn Herndon
University of Central Florida

Recommended Citation
https://stars.library.ucf.edu/honorstheses1990-2015/1836
UNDERSTANDING THE ROLE OF A HEMERYTHRIN-LIKE PROTEIN IN

MYCOBACTERIUM TUBERCULOSIS

by

CAITLYN L. HERNDON

A thesis submitted in partial fulfillment of the requirements for the Honors in the Major Program in Biotechnology in the Burnett School of Biomedical Sciences and in The Burnett Honors College at the University of Central Florida
Orlando, Florida

Summer Term 2014

Thesis Chair: Dr. Kyle Rohde
ABSTRACT

According to the Centers for Disease Control and Prevention (CDC), 8 million people each year are infected with *Mycobacterium tuberculosis* (*Mtb*) leading to 1.5 million deaths annually. This staggering number calls for advancements in understanding this bacterium so progress can be made in treating and preventing the disease. It is particularly important to understand mechanisms by which TB survives inside hostile host immune cells known as macrophages and within hypoxic granuloma lesions of the lung. Preliminary microarray data has shown that a TB gene known as Rv2633c is induced upon macrophage invasion. Bioinformatic analysis of Rv2633c coding sequence shows the product of Rv2633c has homology with hemerythrin-like proteins. Hemerythrins are a class of proteins commonly used to bind oxygen and sense nitric oxide and iron, leading us to hypothesize a role for Rv2633c in surviving hypoxic or nitrosative stress encountered within macrophages and granulomas. My first aim will be to generate a reporter strain of *Mycobacterium smegmatis* (*Msm*) expressing the mCherry fluorescent protein driven by the Rv2633c promoter. This tool will allow us to determine the stress conditions (i.e. hypoxia, nitric oxide treatment, acid pH) that activate expression of this gene by measuring the change in fluorescence. Linking the regulation of Rv2633c to specific environmental cues relevant to infections *in vivo* will provide insight into the role of this unique protein. Secondly, a knockout mutant of Rv2633c in the attenuated *M. bovis* BCG will be constructed and characterized to determine the importance and function of this protein during TB infections.
ACKNOWLEDGEMENTS

I would like to express my deepest thanks to all of the wonderful people who have helped me reach this goal with their kind words and encouragement. First, I would like to acknowledge my mentor and Thesis Chair, Dr. Kyle Rohde for the many hours he spent revising my thesis and helping me to understand my project’s full potential. His guidance has helped develop me as a researcher and I am very thankful to have Dr. Rohde as a mentor.

I would also like to thank all the members of the Rohde Lab for creating a work environment that is uplifting and positive. A special thank you goes to Sandy Geden for helping me countless times along the way. She is dedicated to solving problems and helping anyone who is hitting a rut.

Lastly, I would like to thank my friends and family for the many hours of encouragement. Your support means the world to me and I could not have had a better support system.
# TABLE OF CONTENTS

CHAPTER 1: SPECIFIC AIMS

AIM 1: To characterize the function and genetic organization of Rv2633c using bioinformatics and transcript analysis.

AIM 2: To characterize the transcriptional regulation of Rv2633c using fluorescent reporter bacteria strains.

AIM 3: Use reverse genetics in Mycobacterium bovis BCG (BCG) to determine the phenotypic effect of deleting Rv2633c.

CHAPTER 2: BACKGROUND

The Disease

Evolutionary History

Mycobacterium tuberculosis Pathogenesis

Microarray

CHAPTER 3: MATERIALS AND METHODS

Bioinformatics Investigation

Cell Culture

Polymerase Chain Reaction (PCR)

FastCloning

Rationale:

Method:

Creating the Promoter Reporter Strain

Rationale:

Method:

Characterizing the Promoter

Rationale:

Method:

Generating an Rv2633c Knockout Mutant
Rationale: .........................................................................................................................25
Method: ..............................................................................................................................25
Reverse Transcriptase PCR (RT-PCR) ..................................................................................29
  Rationale: ............................................................................................................................29
  Method: ...............................................................................................................................29
Quantitative Reverse Transcriptase PCR (qRT-PCR) .............................................................30
  Rationale: ............................................................................................................................30
  Method: ...............................................................................................................................31
CHAPTER 4: RESULTS ........................................................................................................33
  Bioinformatic Investigation ..................................................................................................33
  Creating the Promoter Strain ..............................................................................................35
  Characterizing the Rv2633c Promoter ..............................................................................40
  Creating the Knock Out Strain ............................................................................................41
  Reverse Transcriptase PCR .................................................................................................42
  Quantitative Reverse Transcriptase PCR (RT-PCR) ..........................................................44
CHAPTER 5: DISCUSSION ..................................................................................................46
APPENDIX: PRIMER CATALOG .........................................................................................52
REFERENCES .......................................................................................................................56
LIST OF FIGURES

Figure 1: Evolutionary tree of common Mycobacterium species ........................................7
Figure 2: Microarray Data for Rv2632c and Rv2633c mRNA Levels in PhoPR and WhiB3 Mutants .........................................................................................................................................................................................11
Figure 3: Microarray heat map expression data for Rv2632c and Rv2633c ........................12
Figure 4: Creating the Fluorescent Promoter Reporter Strain ........................................20
Figure 5: Characterizing the Promoter .................................................................................24
Figure 6: Creation of the Rv2633c Knock-Out Strain .....................................................28
Figure 7: BLAST Alignment of Rv2633c and Surrounding Regions with Common Species of Mycobacterium .........................................................................................................................33
Figure 8: Hemerythrin Binding Domain. ............................................................................34
Figure 9: Visual evidence for Rv2633c promoter activity in Msm ..................................35
Figure 10: Fluorescent microscopy of pVV2633cRG strain .............................................36
Figure 11: Plate reader data comparing mCherry and GFP fluorescence of pVV2633cRG and pVVRG .........................................................................................................................................................................................37
Figure 12: Wild type Mtb containing the Rv2633cRG fluorescent promoter construct ..........38
Figure 13: Fluorescent Rv2633cRG promoter strain transformed into wild type Mtb and a PhoP mutant .........................................................................................................................................................................................39
Figure 14: Fluorescence levels of the Rv2633cRG construct in wild type Mtb and a PhoP mutant. .........................................................................................................................................................................................40
Figure 15: Screening for positive clones during each step of generating the Rv2633c knockout.
........................................................................................................................................42

Figure 16: Reverse transcriptase PCR on wild type Mtb cDNA. ........................................43

Figure 17: qRT-PCR performed on wild type Mtb cDNA amplifying Rv2632c and Rv2633c ....45
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC</td>
<td>Centers for Disease Control and Prevention</td>
</tr>
<tr>
<td>Mtb</td>
<td><em>Mycobacterium tuberculosis</em></td>
</tr>
<tr>
<td>BCG</td>
<td><em>Mycobacterium bovis</em> Bacille Calmette-Guerin</td>
</tr>
<tr>
<td>TB</td>
<td>Tuberculosis</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny broth</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>NEB</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complimentary deoxyribonucleic acid</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative reverse transcriptase polymerase chain reaction</td>
</tr>
</tbody>
</table>
CHAPTER 1: SPECIFIC AIMS

According to the Centers for Disease Control and Prevention (CDC), 9 million people each year are infected with Mycobacterium tuberculosis (Mtb). Of those infected, Mtb kills 1.5 million annually (1). This staggering number calls for advancements in understanding this bacterium so progress can be made in treating and preventing the disease. Understanding mechanisms by which Mtb survives inside hostile host immune cells known as macrophages and within hypoxic granuloma lesions of the lung are of particular interest. Preliminary microarray data has shown that a gene of Mtb known as Rv2633c is induced upon macrophage invasion. Bioinformatics analysis of Rv2633c coding sequence shows the product of Rv2633c has homology to hemerythrin-like proteins. Hemerythrins are a class of proteins commonly used to bind oxygen and sense nitric oxide, leading us to hypothesize a role for Rv2633c in surviving hypoxic or nitrosative stress encountered within macrophages (2, 3). Recent analysis of Rv2633c also revealed a previously unidentified Dps (DNA protection during starvation) domain, which serves to protect DNA by protecting DNA under stressful conditions such as hypoxia and starvation that Mtb encounters in vivo during infection (4-6). The major goal of this project is to understand the role of Rv2633c during intracellular survival and persistent infection of Mtb. We hypothesize that Rv2633c may use a hemerythrin-like protein to play a key role in sensing oxidative stress from the environment and produce a Dps-like protein to protect its DNA. By gaining additional knowledge into these crucial mechanisms, it will be possible to better treat this disease for years to come.
**AIM 1: To characterize the function and genetic organization of Rv2633c using bioinformatics and transcript analysis.**

We will use bioinformatics tools to analyze the Rv2633c genomic region in order to identify potential functional domains, homologs, and promoter elements. In addition, RT-PCR will be conducted to see if Rv2633c is co-transcribed with the adjacent gene Rv2632c. If the data supports a linkage of the two, subsequent studies of Rv2633c would be designed to take the possible involvement of Rv2632c into account.

**AIM 2: To characterize the transcriptional regulation of Rv2633c using fluorescent reporter bacteria strains**

By fusing the fluorescent protein mCherry downstream of the promoter region of Rv2633c, transcription levels can be easily measured under a variety of conditions. Additionally, the fluorescent reporter strain will be transformed into wild type *Mtb*, a strain with a PhoP mutation, and a strain with a WhiB3 mutation to confirm that these transcription factors regulate Rv2633c. Fluorescence would be significantly lower in the mutant strain lacking the transcription factor than the wild type if the gene is regulated by that protein. Stress inductions will serve to identify environmental cues that activate this promoter. Based on the homology of Rv2633c to hemerythrins and Dps, we will focus on the response of the bacteria to hypoxia, nitric oxide, and stationary phase. Sequences crucial for promoter function will also be identified by mutagenesis and measurement of fluorescence.
AIM 3: Use reverse genetics in Mycobacterium bovis BCG (BCG) to determine the phenotypic effect of deleting Rv2633c.

A knockout of Rv2633c will be created in an attenuated surrogate for Mtb, BCG, using homologous recombination by replacing the gene with a hygromycin marker. In parallel, senior members of the lab will then transform the plasmid into wild type Mtb and observe the effects under stresses discussed in Aim 2.
CHAPTER 2: BACKGROUND

The Disease

Although any part of the body can be infected with TB, pulmonary TB is the most common manifestation (23). Patients who have this disease develop a severe cough which can even include coughing up blood, night sweats, and fever (23). Tuberculosis is transmitted through the air by an infected individual coughing, spitting, speaking, or doing any other action which could make the bacteria airborne (23). Often the disease will lay latent in the body for a long period of time, frequently for a number of years, before the patient shows any symptoms. This makes TB particularly dangerous, as one-third of those infected with the disease are latently infected. Though an infected individual only has a 5-10% chance of developing TB once they’re been exposed to it, the transmission process of TB is highly efficient. It can take only a few $Mtb$ cells to infect another individual (3).

The fact that tuberculosis is a deadly disease is no new development in the world. Samples of spinal fluid in Egyptian mummies dated from 2400 BC have shown the presence of tuberculosis in a number of victims (7). Hippocrates even noted in 460 BC that physicians should not visit patients who had TB (although named “phthsis” at the time), because their assured death would ruin the doctor’s reputation (7). It can be seen that TB has persisted for a number of years, showing a slight decline around the discovery of penicillin, but increasing in the number of cases in the 1980’s. Since then, the number of cases of TB has been on the decline, although still highly prevalent (8).
Tuberculosis will kill approximately 3 people every minute, and yet the most popular treatment for the disease is outdated and requires close supervision. Treatment for both latent and active TB requires the patient to take a cocktail of drugs for an extended period of time. Following an initial dosage phase of two months, the patient must then continue to take many antibiotics for an additional five to seven months (25). Frontline drugs for treating drug susceptible TB include isoniazid, rifampin, ethambutol, and pyrazinamide. TB treatments must be taken carefully, because taking the drugs incorrectly or not finishing the treatment can cause the TB still alive to become drug resistant, which as a result is much harder and expensive to treat (25). Unfortunately, many strains of Mtb have become drug resistant to many of these common treatments. Drug resistant strains require second-line drugs that have a longer duration period and more side effects than the frontline treatments (23, 25). Future research on Mtb must focus on mechanisms that can be inhibited to better fight the disease.

Additionally, a vaccine for Mtb is available consisting of an attenuated strain of M. bovis known as BCG. However, Mtb vaccines given to children are often not effective in adults (1, 9, 23). In many foreign countries receiving the vaccine is common, especially in areas where HIV and other diseases which attack the immune system are prevalent. Those who work in environments where tuberculosis is common may receive the vaccine as well. While BCG has proven effective for preventing TB in infants, its efficacy in adolescents and young adults has not been successful in preventing pulmonary TB (1). Many citizens born within the United States have not received this vaccine due to its low efficacy rate and the low probability of being infected with Mtb in the country. BCG also interferes with tuberculin tests, screens which
monitor if someone has developed tuberculosis. This causes false positives in patients, making it difficult to determine where a tuberculosis infection may be in population (25). Due to the overall lack of efficacy of the BCG vaccine to prevent the form of TB most responsible for disease transmission, it is clear that this vaccine is not acceptable for the world-wide demand of an effective TB vaccine.

The global TB crisis stems from the fact that modern drugs, vaccines, and diagnostic techniques used to fight the disease are not adequate to effectively control the disease. For this reason, WHO developed the six point “Stop MTB” strategy which includes building awareness of the drug and getting vaccines to rural areas. Emphasis is currently placed on funding research to generate new drugs and vaccines, and to improve upon treatments currently available (23). These goals were set to be reached by 2015, and WHO has fallen drastically behind.

**Evolutionary History**

The *Mycobacterium tuberculosis* complex (MTC) is a group of closely related mycobacteria (>99.5% identical) that cause TB of some sort (10). Figure 1 shows an evolutionary tree of members of the MTC and their relation to other mycobacteria. From this diagram it can be seen that species of mycobacterium most closely related to *Mtb* (which form the MTC) display the most similar symptoms of the disease, while increasing evolutionary distance from the MTC contains bacteria that are gradually less pathogenic. *M. bovis* commonly infect cows, and *M. tuberculosis* infects humans. *M. ulcerans* and *M. marinum* can create
lesions on the skin of humans, but *M. marinum* is obtained from aquatic sources (11). One step further shows species which are not as pathogenic; *M. paratuberculosis* can commonly be found in the environment and are usually not harmful to humans unless they are severely immunocompromised. Finally, fast-growing nonpathogenic mycobacteria like *Msm* are the most evolutionarily distant species.

It was once thought that *Mtb* may have been transmitted to humans from cattle around 9,000 years ago, but it can be seen that this is not the case. More recent research suggests that *M. tuberculosis* preceded *M. bovis*, and that humans may have actually infected cattle first with the disease which then diverged into a species specific agent. The competing view holds that both *M. tuberculosis* and *M. bovis* diverged at similar times from a founder strain (3, 12).

![Evolutionary tree of common Mycobacterium species](image)

**Figure 1: Evolutionary tree of common Mycobacterium species.** The evolutionary tree depicts that species such as *M. tuberculosis* and *M. bovis* that cause tuberculosis in humans and cattle respectively, are more closely related than species like *M. ulcerans* which causes severe blister-like skin lesions or *Msm* which does not cause disease.
Conservation of a genetic region typically indicates an important part of the genome. By comparing genomes across species, it is possible to predict which conserved regions are likely vital to pathogenesis by relating it to the species that have gained or lost the sequence. For example, a conserved sequence among virulent strains of a pathogen that has been modified or lost in non-virulent relatives may indicate a mechanism by which the strains are virulent. Likewise, a gene conserved only in human pathogens may indicate a mechanism by which the pathogen survives in the human body. This type of assumption is based on one of Koch’s molecular postulates, “the phenotype or property under investigation should be associated with pathogenic members of a genus or pathogenic strains of a species”. Keeping this in mind, it is important to investigate members of the Mycobacteria family in order to determine what commonalities are present between branches of the family (13). This will help to determine what sort of strategies Mtb adopts to so efficiently infect humans.

Since Mtb must be contained in a BSL 3 Laboratory, safer surrogate organisms are often used to study mycobacterial genetics and virulence factors. BCG (Bacille Calmette-Guerin), the attenuated M. bovis used in TB vaccines, is >99% identical to TB (10). This high level of similarity to the pathogenic strains make it a good substitute for conducting experiments in a BSL2 environment. During the course of this investigation, Msm will also be used as a surrogate organism since it is fast growing unlike BCG which can take weeks to form a colony.
Mycobacterium tuberculosis Pathogenesis

One striking characteristic of TB pathogenesis is its ability to survive within its human host for an extended amount of time (14). The pathogen has evolved particular mechanisms which help it to survive within the conditions presented by the host cell, but little is known about these adaptations. The ability to survive and replicate within macrophages remains one of TB’s best strategies for surviving within a host for many years. Macrophages are large phagocytic cells that are members of the immune system, and are recruited to sites of infection to eradicate any foreign cells. TB has acquired the ability to not only survive within the nutrient-lacking environment of these cells, but is also able to replicate inside of them (15). It is within these cells that TB spends much of its infectious cycle within humans. The conditions within macrophages are much different than the environment around it, encountering stresses including antimicrobial peptides, low pH, hypoxia, reactive oxygen and reactive nitrogen species, as well as lipases, proteases, and DNAases. TB must also have the ability to quickly adapt between the many differing environments (3, 15).

In order to understand this relationship between macrophages and Mtb, it is important to look at how the environment is formed. Initially, Mtb bacilli are inhaled into the lungs where the bacterial cell is phagocytized by macrophages residing in the local alveolar tissue. Here, an inflammatory response is induced, which brings foamy macrophages, mononuclear cells, and T lymphocytes to the infection site. The outside of the conglomeration is then coated with collagen and other fibrous material that forms a hardened surface to the structure. These cells form a barrier around the infected macrophages, and the lesion as a whole is called a
granuloma (or tubercle, where the disease gets its name), which is the body’s attempt at localizing the infection (3).

Approximately 9% of TB cases are asymptomatic because the immune response has successfully localized infection. However, when the integrity of the immune system fails the granuloma will collapse in a process known as cavitation, disseminating many viable \textit{Mtb} cells into the host. Reasons for immune system changes could be the result of old age, HIV, or stress, which are all common for many people infected with tuberculosis (23). The release of the granuloma contents produces a severe cough that is one of the symptoms for tuberculosis. Coughing also causes the spread of the disease through the air (3).

It is the center of this formed granuloma and macrophages that has been shown to be hypoxic. Many bacteria would die under such conditions, but \textit{Mtb} can persist and replicate in such low levels of oxygen which indicates specific mechanisms for sensing the environment and adapting as necessary must be present (15). Transcription factors are sets of genes that help a pathogen establish its virulence inside a host. Two transcription factors which have been proven to be required for \textit{Mtb} in both macrophages and \textit{in vivo} are PhoPR and WhiB3 (16, 17, 26). Since these transcriptional regulators are activated in response to hypoxia, acidic pH, and oxidative stress, the genes they regulate must have some key feature in the ability for the pathogen to survive. Understanding the roles of genes influenced by these transcription factors will help to see what mechanisms \textit{Mtb} uses to survive these conditions.
Microarray

Microarray experiments performed by Dr. Kyle Rohde suggested a role for a gene called Rv2633c during macrophage survival and implies that the gene may be regulated by both PhoPR and WhiB3. When invading and growing within macrophages, mRNA levels of Rv2633c were much higher than when the cell is not exposed to stress (14).

Relative mRNA levels during macrophage infection of Rv2633c in wild type *Mtb* in ratio to a ΔphoPR and ΔwhiB3 mutant are displayed in Figure 2. Expression of Rv2633c and its adjacent gene Rv2632c are markedly lower in mutant strains of *Mtb* than in the wild type strain, suggesting the genes may be regulated by these transcription factors.

![Figure 2: Microarray Data for Rv2632c and Rv2633c mRNA Levels in PhoPR and WhiB3 Mutants.](image)

Levels of Rv2632c and Rv2633c can be seen to fluctuate in similar patterns in the PhoPR and WhiB3 mutants. This may indicate possible existence on the same operon.
Levels of Rv2632c and Rv2633c exhibit similar expression profiles in mutant strains (Figure 2), indicating they may possibly be located on the same operon. Both genes are not activated two hours after macrophage infection, but expression is very high at days 2 and 4, tapering off until day 12 (Figure 3). If the two gene being on an operon were the case, it would be important to take into account the role of Rv2632c when determining the role of Rv2633c. Being transcribed on the same operon may link the genes to similar functions, or functions that would be sensed and activated by common environmental cues.

Figure 3: Microarray heat map expression data for Rv2632c and Rv2633c. The data showed that that Rv2633c and its adjacent gene, Rv2632c, have similar expression profiles during macrophages infection. The two mRNA transcripts are most abundant around day 2 and expression level drops significantly around days 10 and 12.
CHAPTER 3: MATERIALS AND METHODS

Bioinformatics Investigation

Analysis of Rv2633c was done in silico using existing databases that compared the sequence and structure of Rv2633c to homologs in similar bacteria. Using sites such as tbdb.org and xbase.org, the Rv2633c region was compared in both pathogenic and nonpathogenic species of mycobacterium. A BLAST (basic local alignment sequence tool) search was also utilized to look for sequence alignments and conclude any potential structures of the resulting protein. Lastly, array data in these databases was used to identify and analyze expression patterns of the genes of interest.

Cell Culture

Throughout this project, several different strains of bacteria will be used that all grow and thrive optimally in different environments. The amount of time it takes to obtain a culture of a certain bacteria will depend on what is being studied, as some species take longer to grow than others. Below is an overview of what conditions were used for each species.

*Escherichia coli* (*E. coli*) cultures are typically grown on agarose lysogeny broth (LB) plates with antibiotic added to the plate. This will depend on what plasmid is being studied as selectable markers will vary. *E. coli* is typically grown in overnight cultures of 10 ml with 50 µL of 50 mg/ml antibiotic added. Cultures are shaken overnight at 250 rpm at 37°C.
*Msm* was grown in LB broth with Tween added. Since the cell wall of *Msm* is waxy, cells will clump together and the addition of Tween detergent prevents this from happening. *Msm* uses antibiotics like *E. coli* does, but are grown for about two days at 37°C, shaken slowly at ~50 rpm.

**Polymerase Chain Reaction (PCR)**

Polymerase chain reaction (PCR) provides a way to amplify a DNA fragment of interest exponentially using DNA polymerase and primers to target the sequence of interest. Primers are short strands of DNA which are designed to flank the DNA fragment which will be amplified. The DNA template contains the desired amplified piece. Forward and reverse primers anneal to a specific region of the template and polymerase binds to these regions, synthesizing new strands complementary to the template using deoxynucleotide triphosphates (dNTPs). The reaction is then repeatedly heated and cooled in a thermal cycler, causing the fragments to multiply many times.

A typical PCR reaction contains 16.5 µl de-ionized water, 5 µl 10x GC buffer, 1 µl DNA template, 1 µl forward primer, 1 µl reverse primer, 0.5 µl deoxynucleotide triphosphates (dNTPs), and 0.25 µl Phusion polymerase for a reaction total of 25 µl. A standard PCR cycle is shown below:
98°C- 30 seconds

98°C- 10 seconds

x°C- 30 seconds  x17

72°C- y seconds

72°C- 5 minutes

25°C- hold

The variable x represents the appropriate annealing temperature for the two primers to hybridize with their target sequence. Based on optimal Phusion parameters, primers are typically designed to work at 65°C using a Tm calculator. This temperature can be calculated using the Tm calculator available online from New England Biolabs by entering the primer sequence into the calculator. The variable y represents the extension time for the reaction. Typically, the time should increase by fifteen seconds for every kilobase of target sequence length. For example, a 2,000 bp piece would have an extension time of 30 seconds.

Once the PCR is complete, the reaction is run on an agarose gel to confirm that there is a band at the correct molecular weight by comparison to a ladder of known band size run parallel to the sample. This verifies that the fragment of interest was properly amplified.
FastCloning

**Rationale:** Recombinant DNA technology provides a way to create new pieces of DNA by combining fragments of different function or origin for a variety of experimental applications. The product is a new DNA construct which can then be expressed, or cloned, into a host cell and the DNA will be present. Generating recombinant DNA products has traditionally been a long and tedious process. Standard methods have relied on restriction enzyme digestion, which requires a specific sequence of DNA to be present for the enzyme to cut. The DNA fragment of interest is then inserted at the restriction site, but this means that the fragment is limited to which restriction sites and enzymes are available. DNA fragments must also be purified on an agarose gel to ensure fragments are the right size and that any contaminating pieces are not present. Ligation is necessary to then seal any gaps in the final DNA construct, which adds yet another step to the process.

A method widely utilized during this project is a polymerase chain reaction (PCR) based cloning method called FastCloning, which is purification, digest, and ligation free (18). Because FastCloning does not use the traditional methods of restriction enzymes but relies on primer design, it is possible to insert a fragment into any position on a plasmid vector, provided the sequence of the insertion site is known.

**Method:** Both the insert and the vector were amplified using PCR. Primers are designed to amplify the vector into a linear piece of DNA, open at the desired position at which to insert the fragment. Once the two reactions were verified on a gel, they were mixed together at a 1:4,
1:1, and 4:1 insert to vector ratio at a total volume of 8 μl. The mixtures then had 2 μl of DpnI added which digested the methylated strands of DNA. DpnI will digest original, methylated plasmid DNA but not the unmethylated insert and vector fragments generated by PCR. This serves to eliminate false-positive clones that would result if the template plasmid is introduced into E. coli during transformation. This allows users to forego the time consuming vector digestion and purification steps traditionally performed in other cloning methods. Once the insert, vector, and DpnI reactions have been incubated at 37°C for one hour, they are directly transformed into chemically competent NEB 10-beta E. coli cells.

Transformation of the FastCloning reactions began by thawing the NEB 10-beta E. coli cells on ice for 10 minutes. 1-5 μl of the FastCloning mix was then added to an aliquot of the thawed cells and placed on ice for 30 minutes, followed by a 42°C heat shock for 30 seconds and placed on ice for 5 minutes. Following addition of 200 μL of SOC, the mix was then shaken at 250 rpm at 37°C for one hour. After incubation, 50-100 μl of the reaction is plated on LB-hygromycin plates.

The plates which are used during the transformation contain antibiotic that selects for the amplified plasmid. In the case of this study, the pVV16 vector has a hygromycin marker, so the transformations were plated on Hyg50 (50 μg/ml) LB plates. Any colonies that appear as a result of the transformation should contain plasmids possessing the desired insert because only then will they possess the antibiotic resistance gene. However, false positives can arise from
incomplete Dpn1 digestion or rare self-ligation of the fragment alone. For this reason, these colonies are then screened using the colony PCR method detailed below.

To screen for positive clones, a small colony from any plate is selected at random and boiled in 20 μl of water for 5 minutes in order to release their DNA. The sample is then centrifuged at 10,000 xg for 5 minutes so that cellular debris will collect at the bottom of the tube. The lysate containing DNA from the colony is then used as the template DNA in a PCR reaction. For colony screening, primers used to amplify the insert will be used. Since the selectable marker ensures that only colonies which have taken up the plasmid can grow on the plate, this step confirms that the insert is present too. The cycle used will be the same as in standard PCR, but with 30 cycles instead of 17.

Once positive colonies have been confirmed, it is inoculated into LB broth plus antibiotic to culture the clones containing the desired plasmid. The plasmid is then purified and the plasmid extracted from the sample using an alkaline lysis protocol from Qiagen, and positively cloned DNA is then obtained. The DNA is then sequenced to verify that mutations are not present.

Creating the Promoter Reporter Strain

Rationale: By engineering fluorescent protein expression to be driven by a specific promoter, we will generate a powerful tool to facilitate the study of gene regulation. Reporter strains allow real-time measurement of promoter activity in the whole cell using a plate reader.
As the promoter of interest is activated, the fluorescent protein downstream of the promoter is transcribed and fluorescence levels of the culture will increase. This reporter strain will also contain a dual color approach, using green fluorescent protein (GFP) and mCherry to detect ratiometric changes.

**Method:** A shuttle vector called pVVRG will be used that has the ability to replicate in both *E. coli* and mycobacteria. This vector contains green fluorescent protein (GFP) controlled by a strong *hsp60* promoter and mCherry downstream of a second strong promoter called *smyc*. This shuttle vector also features kanamycin and hygromycin resistance.

Primers were designed to amplify a ~500 bp region upstream of Rv2633c containing the promoter from *Mtb* chromosomal DNA (1 µl chromosomal TB DNA at 20 ng/µl, 1 µl primer 1: 2633pro_mCh_F, 1 µl primer 2: 2633pro_mCh_R, 0.25 µl dNTP, 5 µl GC Buffer, 0.25 µl Phusion High Fidelity DNA Polymerase, 16.50 µl De-ionized water, for a 50 µl total; primers work at a 67°C annealing temperature with a 30 second extension time). The pVVRG plasmid was also amplified (1 µl chromosomal TB DNA, 1 µl primer 1: pVVmCh.FC_F, 1 µl primer 2: pVVmCH.FC.Rpro, 0.25 µl dNTP, 5 µl GC Buffer, 0.25 µl Phusion High Fidelity DNA Polymerase, 16.50 µl De-ionized water; primers work at a 65°C annealing temperature with a 1.5 minute extension time).

FastCloning will allow the *smyc* promoter of pVVRG to be replaced by the Rv2633c promoter via homologous recombination. The product of this transformation is then PCR screened for positive clones, and confirmed by sequencing for accuracy [5].
Once this sequence has proven to contain no mutation, it was electroporated into Msm. This was required because the cellular membrane of Msm is particularly thick. As Figure 4 shows, the result of this transformation was a pVVRG derivative with the Rv2633c promoter driving the mCherry promoter. By measuring changes in the ratio of mCherry fluorescence to baseline GFP fluorescence (a control), the activity of the Rv2633c promoter could be monitored.

Figure 4: Creating the Fluorescent Promoter Reporter Strain. The promoter reporter strain was created by first amplifying the Rv2633c promoter region in front of the Rv2633c gene (A). A preexisting plasmid known as pVVRG contained a strong smyc promoter in front of a sequence coding for the fluorescent protein mCherry. This plasmid also features a GFP sequence paired with another promoter for the purpose of a control (B). FastCloning of the Rv2633c promoter region and pVVRG replaced the smyc promoter with the promoter of Rv2633c, and the activity of the promoter can then be monitored by mCherry levels (C).

The fluorescence level of GFP will be consistent throughout any condition the bacteria is exposed to because the strong, hsp60 promoter will not be affected by the cell’s environment. Since the Rv2633c promoter will respond to the environmental stresses and activate mCherry
expression, the ratio of mCherry to GFP fluorescence will indicate the relative induction of the Rv2633c promoter.

In addition to plate reading, the colonies containing the pVVRG2633c construct were observed underneath a fluorescent microscope (EVOS FL). A positive colony from the Msm plates was lightly streaked onto a glass slide and fixed into place with a cover slip. Images were captures using both Texas Red and GFP filters to visualize mCherry and GFP signals produced by pVVRG2633c Msm.

In the future, we plan to see if Rv2633c is induced by hypoxia stress by incubating the reporter strain under normal oxygen levels and hypoxia (~1-5 % O2) in a hypoxia chamber. The ratio between GFP and mCherry will indicate if Rv2633c is responsive to hypoxia; an increase in mCherry fluorescence indicates induction. Additionally, links to PhoP indicate a relation to acidic conditions. A gradient of pH with liquid media buffered to a variety of pH levels (~pH 5-7) will be used to determine if the Rv2633c is upregulated when these conditions are present.

Once the construct was confirmed to work in Msm, the construct was transformed into Mtb and BCG for further experimental purposes. Since Msm does not have an Rv2633c homologue, using the construct in Mtb and BCG may more faithfully mimic the “real life” situation. Since Mtb must be handled in a BSL3 laboratory, this part of the project was carried out by a different member of the team who has access to the laboratory. However, BCG could be used for additional experiments in a BLS2 setting, and then confirmed in Mtb. All strains will
be exposed to various environmental conditions and monitored using the fluorescent strain as a reporter of activity.

Characterizing the Promoter

Rationale: In addition, our reporter system can be used for detailed characterization of the Rv2633c promoter itself. PCR-based mutagenesis will be utilized to truncate the ~500 bp region which is assumed to contain the true promoter. Fragments of this region will be deleted while mCherry will be measured until the fluorescence drops, indicating the loss of genetic elements essential for either Rv2633c basal expression or stress induction by PhoP and WhiB3. This will allow us to characterize the sequence required for basal activity of the gene, and the Rv2633c promoter will be defined.

The initial promoter region cloned upstream of mCherry included ~500 bp that was chosen because most promoters are found within this range. The true promoter was somewhere in this region, but not fully defined. An approach called Round-the-horn PCR (RTH-PCR) (19), which can delete small fragments of DNA or perform point mutations, was utilized to modify the Rv2633c promoter region in the Rv2633c:mCherry reporter construct in order to find the true promoter region. Additional shorter fragments of this region were also prepared to be tested.

This approach involves designing a forward primer which hybridizes at a position defining the downstream fragment, and a reverse primer that encodes for the upstream part of
the region that will be kept. This puts the deleted fragment of the gene between the two regions which the primers amplify. These primers are phosphorylated prior to amplification. Thus, when these primers amplify the fluorescent reporter plasmid, only the part of the vector which contains the desired sequence is made and the fragment which is being deleted is not present. The PCR product is then ligated together using the phosphorylated ends of the primer, as Figure 5 shows. This new, truncated plasmid is then transformed into *E. coli* and expression of mCherry is monitored with a plate reader.

**Method:** The forward and reverse primers were first individually phosphorylated at 37°C for one hour (37.5 μl de-ionized water, 5 μl 10x kinase buffer, 1 μl 100 mM ATP, 0.5 μl 100 mM MgSO₄, 5 μl 10 μM primer, 1 μl T4 Polynucleotide Kinase (PNK)). They were then incubated at 65°C for 20 minutes to heat inactivate the T4 PNK.

Next, the phosphorylated primers were utilized in a PCR to amplify the desired section of a plasmid while omitting the area between the primers (Forward primer: 2633pro_mCh_F2, Reverse primer: pVVmCh_FC-Rpro). Once the PCR reaction is complete, the product was run on an agarose gel to ensure that a fragment of the correct size was obtained. This fragment will be the size of the original plasmid minus the amount deleted.

When the fragment appeared to be the correct size, 2 μl of *DpnI* was then added to the mix and incubated at 37°C for 2 hours. Once the incubation was complete, the PCR product was ligated together using the phosphorylated ends of the primers to re-circularize the plasmid (1.5 μl Fast-link ligation buffer, 0.75 μl 10 mM ATP, 3 μl PCR product, 1 μl Fast-link DNA ligase, 8.75
μl de-ionized water; incubated at room temperature for 2 hours, then heat inactivated the Fast-link DNA ligase at 70°C for 15 minutes). This time when the ends are joined together, the plasmid did not contain the deleted section. After the final incubation is complete, the reaction mix was transformed into competent *E. coli* cells.

**Figure 5: Characterizing the Promoter.** (A) Characterizing the promoter utilizes a method known as Round the Horn PCR. First, phosphorylated primers are used to amplify the pVV16 plasmid containing the promoter region of Rv2633c, but a small area which will be deleted is omitted from amplification. (B) Once the PCR is complete, the promoter region does not contain a certain region of the original sequence and the ends of the fragment are also phosphorylated. (C) The phosphorylated ends of the fragment are ligated together, re-circularizing the plasmid which now has a truncated promoter region.

Using a plate reader, fluorescence will indicate which constructs contained the promoter region. Fluorescence which matches that of the ~500 bp region will be from colonies that have retained activity. Partial loss of fluorescence would indicate that a promoter is still present, but has some transcription factor binding sites missing. A loss of fluorescence will mean the baseline promoter region has been truncated and lost. At this point it will be known
that the true promoter has been disturbed and point mutations will be made to define specific base pairs required for wild-type promoter activity.

---

**Generating an Rv2633c Knockout Mutant**

**Rationale:** Reverse genetics provides a way to observe the phenotype of an organism when a particular gene is missing. Using homologous recombination, the gene of interest will be replaced with a hygromycin marker while the areas adjacent to the gene will be left intact. By applying this approach, it is possible to understand the importance of a gene of interest by observing what happens to an organism when it is not present.

**Method:** This process includes cloning regions of the DNA that flank the gene targeted for deletion on either side of a hygromycin selectable marker in the pFCKO knockout vector created in the Rohde lab. FastCloning as described in detail above will be used in order to create this knockout plasmid.

The vector used for this method, pFCKO, contains many unique features. The first of these is a Hyg marker flanked on either side by loxP sites, which allow removal of the marker by Cre recombinases. This is favorable because selectable markers in *Mtb* are limited and removal of the Hyg marker will allow reuse of this marker. As mentioned above, ~1 kb regions on either side of the gene of interest are inserted upstream and downstream of the Hyg marker using standard FastCloning, as Figure 6 shows. Once this construct has been prepared, it is ready for transformation.
A double crossover event will result in the knockout construct replacing the wild type gene. However, single crossover events which integrate the whole plasmid into the chromosome are an unwanted possibility. The *rpsL* gene is a counterselectable marker on the pFCKO vector that allows us to filter out the single cross over event. Bacteria which undergo single crossover and take up the *rpsL* gene are sensitive to streptomycin and they will not grow on a plate containing strep.

To amplify a ~1 kb fragment directly upstream of Rv2633c, forward primer 2633c_Up_F and reverse primer 2633c_Up_R were used in a PCR including *Mtb* chromosomal DNA as the template. Next, the pFCKO vector was amplified using primers that linearized the plasmid at a point directly upstream of the Hyg marker (forward primer: pFCKO_Up_F and reverse primer: pFCKO_Up_R). Standard FastCloning was used to insert the upstream fragment into the plasmid at the correct point. Once colonies were obtained on a plate, screening occurred using the same primers used to amplify the upstream fragment to ensure that the insert was successfully in the plasmid. This clone was name pFCKO_2633Up, meaning that the pFCKO vector had the 2633c upstream region cloned into it.

A similar process occurred for the downstream fragment. A downstream region of ~1 kb directly adjacent of Rv2633c was amplified using forward primer 2633_Dn_F and reverse primer 2633_Dn_R. The pFCKO_2633Up plasmid was amplified using primers that opened the DNA at the region directly downstream of the hyg marker (forward primer: pFCKO_Dn_F and reverse primer pFCKO_Dn_R). Once again, FastCloning was utilized to combine the Rv2633c
downstream fragment and the opened pFCKO_2633Up DNA. Once colonies were grown, the
downstream fragment primers listed above were used to screen the colonies to ensure that the
downstream fragment was present. If these screens indicated the fragment was present, the
construct was taken into consideration to be the final knock out, but named pFCKO_2633c until
a final screening method could verify the knockout plasmid construction was a success.

An additional screening step was used to ensure all three essential element of the
pFCKO vector are present in the construct. Two primers called M13R and M13F that flank the
cutout construct were used to amplify across pFCKO_2633c. Since the upstream fragment,
Hyg marker, and downstream fragments were all ~1kb individually, colonies containing the
desired construct were expected to produce bands at ~3,000 bp. If bands were seen at ~2,000
bp or ~1,000 bp, it would indicate that one or more of these elements were missing.

This plasmid can then be electroporated into mycobacteria where it will cross into the
chromosome by homologous recombination, effectively deleting Rv2633c from the
chromosome. Positive knock-out clones can be identified by their ability to grow on agar plates
containing hygromycin (50 μg/ml) and streptomycin (100 μg/ml). The knock out will then be
assessed by its growth and survival under in vivo stresses such as hypoxia, nitric oxide, acidic
pH, and in macrophages.
Figure 6: Creation of the Rv2633c Knock-Out Strain. (A) Creating the knock out was done in a two-step process. First, the regions 1,000 bp upstream and 1,000 bp downstream of Rv2633c were amplified (B) and cloned on either side of a Hyg marker using two separate FastCloning steps. (C) This construct was the transformed into Mtb, (D) where a double cross over event caused the construct to be incorporated into the genome.
Reverse Transcription PCR (RT-PCR)

**Rationale:** Bacterial genes are often arranged in operons, which include multiple genes transcribed from a single promoter on a single mRNA transcript. Frequently, genes located on the same operon are co-regulated meaning they may have similar function. RT-PCR will be conducted to see if Rv2633c is co-transcribed with the adjacent gene Rv2632c. If the data supports a linkage of the two, subsequent studies of Rv2633c would be designed to take the possible involvement of Rv2632c into account.

To determine whether two genes are transcribed into a single mRNA, Mtb mRNA can be transcribed back into DNA using an enzyme known as reverse transcriptase (RT). The cDNA will contain genes in their respective operons since this DNA will be obtained from mRNA present in the bacteria. In order to determine if Rv2632c and Rv2633c are located on the same operon, primers on either side of the gap between the genes will be used to determine if the two genes are transcribed together. If Rv2632c and Rv2633c are on the same transcript, then primers designed to amplify the intergenic region between them will generate a product. If they are not on the same transcript, the primers will be targeting two separate cDNA pieces and the fragment will fail to amplify. As a result, there would be no bands on the agarose gel.

**Method:** During RT-PCR, *MtbcRNA was reverse transcribed into complimentary DNA (cDNA) by reverse transcriptase enzyme (4 μl iScript reaction mix, 1 μL iScript reverse transcriptase, 1 μl RNA (100-250 ng/μl), 14 μL Nuclease-free water; Reaction times may be carried out in a thermocycler: 25°C- 5 minutes, 42°C- 30 minutes, 85°C- 5 minutes, store at 4°C).**
Once the cDNA is obtained, it was used as a template for a standard PCR with primers that flank the gap between the two genes (Forward primer: Rv2632/33_F, Reverse primer: Rv2632/33_R). The PCR was then imaged on a 1% agarose gel.

Two controls will be used for this experiment. In one, RNA would be added to the PCR reaction without adding any reverse transcriptase (4 μl buffer, 1 μl RNA (100-250 ng/μl), 15 μl nuclease-free water; the same thermocycler setting as the reverse transcriptase samples should be used). This “No RT” control is to ensure that the RNA is not contaminated with any DNA products. Since reverse transcriptase is not added to the RNA, no DNA should be present, and no amplified bands would be expected. The positive control involves using the primers for the RT-PCR (Forward primer: Rv2632/33_F, Reverse primer: Rv2632/33_R) on Mtb chromosomal DNA. By adding this PCR reaction, it can be confirmed that PCR conditions were appropriate for amplifying target RNA.

**Quantitative Reverse Transcriptase PCR (qRT-PCR)**

**Rationale:** As a method to quantify the mRNA levels of Rv2632c and Rv2633c in order to determine their co-regulation, qRT-PCR will be utilized. Similar to the RT-PCR, qRT-PCR depends on the reverse transcriptase enzyme in order to create cDNA out of *Mtb* RNA (using the same procedure as RT-PCR to obtain cDNA). The qRT-PCR machine will use this cDNA as a template for a PCR reaction which can quantitatively monitor DNA synthesis in real-time using SYBR green, a fluorescent dye which intercalates into amplified products. Thus, detection of an
increase in fluorescence can be used to quantitate relative amount of mRNA transcripts. By doing this, it is possible to determine if the levels of Rv2633c and Rv2632c are expressed at similar levels and up-regulated under similar conditions, as would be expected if they were in an operon. Samples from macrophage invasion will be used for this quantification in order to see at what time period during infection Rv2633c and Rv2632c are transcribed. The PhoP and WhiB3 mutants will also have their RNA extracted and quantified to determine how these genes are regulated when their hypothesized transcription factors are mutated.

**Method:** Once cDNA is obtained, it is added to a Supermix which contains buffer, dNTPs, polymerase, and SYBR green (12.5 μl iScript Universal SYBR Green Supermix, 2.5 μl Forward primer, 2.5 μl Reverse primer, 5.5 μl de-ionized water). For this experiment, all reactions were done in triplicate to ensure accuracy. Reaction 1 amplified Rv2632c and contained forward primer Rv2632_F and reverse primer Rv2632_R). Reaction 2 amplified Rv2633c and contained forward primer Rv2633c_F and reverse primer Rv2633c_R.

The last reaction was a control which amplified the sigA gene, which is vital for life in the bacteria (Forward primer: SigA_F, Reverse primer: SigA_R). This gene is transcribed in steady, constitutive levels in the cell, and is therefore used as a way to normalize the transcript levels of other genes by measuring their rate of amplification in comparison to this mRNA transcript.

These reactions are all set up in a 96 well clear plate and read in a 7900AT Fast RT-PCR Machine (Applied Biosystems). The PCR cycles are set to:
95°C - 2 minutes

98°C - 10 seconds ← X40

85°C - 30 seconds ←

Since *sigA* is needed in high quantities in the cell, it can be expected that its levels of mRNA transcripts will be seen at about 21 cycles. To ensure the PCR worked, *sigA* is used as a control to make sure this signal is seen around its typical amplification cycle. This signal can then be used to normalize the rest of the data collected by calculating it with the ∆∆Ct method (20). This method determines relative changes in mRNA levels by comparing samples to a known control, normalizing the samples, and then comparing the samples to one another.
CHAPTER 4: RESULTS

Bioinformatic Investigation

Using a basic sequence alignment, it was observed that the sequence coding for Rv2633c is present only in species of mycobacterium which cause harm to humans (Figure 7). In species which infect other animals (such as Mycobacterium marinum, which harms fish) and slow growing species which are non-pathogenic the gene is not present, due either to loss by non-pathogens or gain by pathogens. This data can support the hypothesis that this gene of interest is utilized during pathogenicity in humans.

Figure 7: BLAST Alignment of Rv2633c and Surrounding Regions with Common Species of Mycobacterium. The results of a BLAST search show that species that cause tuberculosis in humans and other animals have retained the Rv2633c gene, where other species of mycobacterium which cause some other type of disease or no disease at all have lost this gene.
The sequence of Rv2633c also revealed the product of the gene to be a hemerythrin. This is a class of protein which can bind oxygen and iron, and the hemerythrin binding domain can be seen in Figure 8. BLAST analysis of the Rv2633c gene also revealed that the C-terminal domain has homology to a protein (~1.8 kDa) known as Dps (DNA-binding protein from starved cells). Dps also shows similarity to ferritins, indicating that the protein may also play a role in ferritin storage.

**Figure 8: Hemerythrin Binding Domain.** (A) The hemerythrin binding domain common to all hemerythrins consists of a bundle of four alpha helices. The domain typically binds iron or oxygen. (B) The hemerythrin binding domain is close to a Dps domain.
Creating the Promoter Strain

A fluorescent reporter strain of the Rv2633c promoter was constructed in order to observe the activity of the promoter region. By doing this, it will be possible to expose the cell to a variety of stresses and use the fluorescent reporter to determine what conditions cause the gene to be activated. In Figure 9, both wild type Msm and Msm with pVV2633RG were plated on the same LB plate. It can be seen when the two cultures grow up side-by-side that the cells with the Rv2633c promoter are visibly more pink than when the species does not have the promoter driving mCherry.

Figure 9: Visual evidence for Rv2633c promoter activity in Msm. Msm pVV2633cRG shows visibly pink colonies (right) versus wild type Msm (left).
When a sample from the *Msm* pVV2633RG reporter strain was viewed under fluorescent microscopy, both mCherry and GFP could be seen. Since both fluorescent reporters were easily detected under a fluorescent microscope, the pVV2633RG construct was a success and a dual-colored fluorescent reporter strain was successfully created (Figure 10). This provides proof of principle for the use of this ratiometric reporter strain coupled with fluorescent microscopy to monitor *Mtb* promoter activation during macrophage infection.

*Figure 10: Fluorescent microscopy of pVV2633RG strain.* Fluorescence microscopy imaging of Msm pV2633RG on an EVOS FL shows both red and green fluorescence, indicating that the construct was a success.
To support fluorescent microscopy, a quantitative measure of fluorescence was obtained using a plate reader. The amount of mCherry and GFP fluorescence was measured for pVV2633RG cultures and pVVRG cultures (in which mCherry expression is driven by smyc promoter). The bar graph displayed in Figure 11 shows the difference between the two strains. It can be seen that pVV2633cRG Msm with the Rv2633c promoter has a much higher level of mCherry than pVVRG Msm, indicating highly active promoter activity.

![Plate reader data comparing mCherry and GFP fluorescence of pVV2633cRG and pVVRG. Plate reader data shows the quantitative difference between the pVVRG fluorescent promoter construct containing the Rv2633c promoter with the same plasmid lacking this promoter in Msm. The Rv2633c promoter appears to be stronger than the smyc promoter.](image)

**Figure 11: Plate reader data comparing mCherry and GFP fluorescence of pVV2633cRG and pVVRG.**
The pVV2633RG fluorescent reporter was then transformed into wild type *Mtb*, which showed bright magenta colonies (Figure 12). This was beneficial because it proved that Rv2633c has high levels of activity in *Mtb*.

![Image](image.png)

**Figure 12:** Wild type *Mtb* containing the Rv2633cRG fluorescent promoter construct. Wild type *Mtb* transformed with pVV2633RG reveals high activity level of Rv2633c promoter.

Figure 13 shows the result of transforming both wild type *Mtb* and a ΔPhoPR mutant with the reporter strain. Wild type *Mtb* once again shows bright pink colonies while the PhoPR mutant is a dull, white color. This result confirms that Rv2633c may be regulated by this transcription factor, and in the absence of PhoPR Rv2633c is not actively transcribed.
Figure 13: Fluorescent Rv2633cRG promoter strain transformed into wild type *Mtb* and a PhoP mutant. Expression of Rv2633c is visibly reduced in the ΔphoPR mutant compared to wild-type *Mtb*.

Cultures from both the wild type *Mtb* and ΔPhoPR mutant *Mtb* were read in a plate reader. The data displayed in Figure 14 shows a significant drop in mCherry levels between the two cultures, with the wild type *Mtb* reporting a much higher level of Rv2633c than the PhoPR mutant. This leads us to conclude that the PhoPR two-component regulator controls Rv2633c.
Characterizing the Rv2633c Promoter

An approach called ‘Round-the-horn, was used to find the baseline sequence needed for promoter activity. In this approach we will delete some of the preexisting ~1 kb region upstream of the Rv2633c promoter and use a plate reader to monitor baseline activity. Currently, primers amplifying the ~500 bp region have been ordered. They were proven to work and optimized using standard PCR. The next step will be to phosphorylate the primers. Once this mutant is transformed into E. coli, it will be possible to tell if the baseline promoter region is located in the ~500 bp sequence by comparing the mCherry fluorescence to that of the ~1kb upstream construct.
Creating the Knock Out Strain

Reverse genetics provides a way to observe the phenol type of a cell when a gene of interest is missing. In order to perform this type of experiment of Rv2633c, a knock out of the gene was constructed. The gel in Figure 15 shows the screening process for obtaining this mutant. In Lane 1, a band at 1,000 bp can be seen indicating only the Hyg marker was present in the plasmid. Lane 2 shows the screening of the intermediate pFCKO_2633cUp construct. Since at this step only the upstream fragment had been inserted into the vector, the band at ~2,000 bp is expected because the primers are amplifying both the upstream fragment and the Hyg marker. In the last lane, the downstream fragment of Rv2633c was amplified and inserted into the pFCKO vector already carrying the upstream fragment. During the final step of the process, pFCKO_2633cKO was amplified and a band can be seen at ~3,000 bp. A band of this size is correct since the upstream, Hyg marker, and downstream fragment are all present. The knockout was considered a success, and this DNA can then be electroporated into mycobacteria.
Figure 15: Screening for positive clones during each step of generating the Rv2633c knockout. It can be assumed that a knockout of Rv2633c was generated since the adjacent upstream and downstream genetic portions appear to be successfully inserted at each step.

Reverse Trascriptase PCR

Since Rv2632c and Rv2633c exhibit similar microarray profiles under macrophage invasion (See Figure 16), it is important to determine if the two genes are located on the same operon. If they are, it is important to study both of the genes since they may have a similar function. Forward and Reverse primers are designed to amplify the intergenic region between Rv2632c and Rv2633c. This product can only be obtained from cDNA if the two genes are co-transcribed on a single mRNA (operon) (Lane 1). For the “no RT” control, reverse transcriptase
enzyme was omitted; a band in this sample would indicate DNA contamination in RNA sample (Lane 2). The primers used to amplify the region were used to amplify the fragment on chromosomal \textit{Mtb} DNA as a control that the PCR worked (Lane 3). This data implies that there was no contamination in the sample since no RNA was present in Lane 2. It can also be seen that PCR was successful since the control in Lane 3 produced a band.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{PCR_result.png}
\caption{Reverse transcriptase PCR on wild type \textit{Mtb} cDNA. Primers amplifying the intergenic region between Rv2632c and Rv2633c did not produce a band, while the control showed that the PCR was working properly.}
\end{figure}
Quantitative Reverse Transcriptase PCR (RT-PCR)

During qRT-PCR, cDNA of wild type *Mtb* was amplified using primers specific for Rv2632c, Rv2633c, and *sigA*. This provided a way to precisely measure the quantity of each transcript present in the sample using SYBR Green fluorescent dye. The qRT-PCR data depicted in Figure 17 shows the results of amplification done with a *sigA* control using wild type *Mtb* RNA. The *sigA* control amplifies at the expected ~21 cycles, indicating the integrity of the template RNA and success of the reactions. Amplification of Rv2632c and Rv2633c transcripts at similar cycles indicated that both of our target genes are expressed at relatively high levels under normal growth conditions. This assay can be used to monitor expression patterns of two genes under different stress conditions to determine if they are coordinately regulated.
Figure 17: qRT-PCR performed on wild type *Mtb* cDNA amplifying Rv2632c and Rv2633c. The graph resulting from qRT-PCR amplifying the genes of interest show that Rv2632c and Rv2633c are amplified in high quantities.
CHAPTER 5: DISCUSSION

*Mycobacterium tuberculosis* is an extremely dangerous disease that infects millions every year. This disease was acknowledged by WHO in the early 1990’s as a global health crisis and since then researchers have begun focusing on understanding the disease better and developing new drugs that will effectively combat the disease (23). Since the development of a TB vaccine is difficult to optimize because of its complex interactions with the immune system, efforts must be made to understand mechanisms by which the disease infects so that drugs treating the disease can be more effective.

With this in mind, our research focuses on understanding a possible method by which *Mtb* can survive in the stressful conditions on macrophages and granulomas of the human body. This environment often has stresses such as hypoxia, hydrogen peroxide, and oxidative stress that kills many infectious cells, but *Mtb* can survive and even replicate in such stresses for many years (15). Here, we present a gene named Rv2633c that encodes a hemerythrin-like protein which may be a helpful mechanism *Mtb* can use during stressful conditions.

Initial bioinformatics analysis of Rv2633c revealed an internal domain with significant homology to a class of proteins known as hemerythrins. Hemerythrins are a class of proteins that have the ability to bind and store oxygen; many marine organisms possess high quantities of hemerythrins, which can help to bind oxygen and help the organism survive in environments where oxygen levels are low (2). Like marine environments, the oxygen levels are low inside of a granuloma. Additionally, hemerythrins have been linked to iron binding, with domains that
bind iron and increased levels of activity in iron-deficient environments (21). The presence of a gene like Rv2633c could indicate a possible mechanism for adapting to such a condition, such as sensing oxygen levels or storing oxygen in hypoxic environments. A previous study found that an E3 ligase linked to an iron-responsive hemerythrin domain played a regulatory role in the stability of the protein. The hemerythrin acted as a switch that mediated the stability of the protein (22). It is possible that the role is similar for a Dps-like protein in Rv2633c.

Rv2633c also includes a C-terminal domain with homology to a protein known as Dps (for DNA protection during Starvation). A previous study found this protein to be present in high levels during the stationary phase of bacterial growth (4). Additionally, Dps is a nonspecific, DNA binding protein that can make a cell more viable in conditions such as increased hydrogen peroxide levels and other oxidative stresses than cells with a Dps mutation (5-6). Interestingly, Dps is actually more stable after binding to DNA. Unbound Dps could withstand temperature of about 65°C before denaturing, whereas Dps bound to DNA withstood temperatures up to 100°C (4). Our data reports a Dps homolog in Mtb, but prior to this study, it was thought that Mtb did not have a Dps homolog (24).

If Rv2633c transcribes a hemerythrin and Dps, it is possible that the hemerythrin region is used to sense the environment and Dps binds DNA for protection in response to an environment with low oxygen or high oxidative stress. This means that this particular gene could be a mechanism by which Mtb protects itself under the harsh conditions of granulomas and macrophages. Like in the E3-ligase example, perhaps the hemerythrin domain regulates the
stability of the Dps domain, and thus a DNA protection mechanism, in response to environmental cues.

To our knowledge, this is the first study to characterize the role and importance of this unique protein for *Mtb* pathogenesis. We used a variety of experimental approaches, including fluorescent reporters, RT-PCR, and reverse genetics to investigate the regulation and function of Rv2633c.

Our first goal was to design a genetic construct to monitor the regulation of the Rv2633c promoter using a fluorescent protein called mCherry. Our data showed that the ~500bp upstream region we cloned did contain elements required for promoter activity in *Msm* and *Mtb*. The construct was remarkably bright in *Mtb*. This construct can now be used to determine when Rv2633c is activated in the cell. After transforming this reporter into a cell, the bacteria can be exposed to various environmental stresses such as hypoxia, nitrosative stress, and varying pH levels to determine how the gene is activated. After exposing the bacteria to such conditions, the fluorescence of the cells will be measured using a plate reader. Those with a higher fluorescence level indicate that the gene was activated, and those with no change mean the gene was not activated. Using this data, it will be possible to see which conditions cause the promoter to be activated by which stresses caused fluorescence to increase. The role of Rv2633c can be implied from this data based on what it responds to in the cell. Fluorescence is expected to increase during hypoxia and low pH due to the hemerythrin like structure of the protein Rv2633c and its apparent regulation by PhoPR and WhiB3.
The transcription factors PhoPR and WhiB3 have been proven to be crucial to *Mtb* survival in macrophages (17). Therefore, it is important to determine if the gene of interest is regulated by either of these factors since that would imply the gene may have a key role in helping the bacteria live under stress. The relationship of Rv2633c to the transcription factor PhoPR can be seen by the way Rv2633c fluorescent levels are significantly lower in PhoPR mutants. This leads us to conclude that Rv2633c is regulated by this transcription factor. This is significant because PhoPR plays a role in TB pathogenesis. Similar experiments will be carried out using a WhiB3 mutant in the lab, to confirm regulation of Rv2633c by this transcription factor.

Due to the similar expression profiles of Rv2632c and Rv2633c in the macrophage infection microarray (Figure 3), it is important to determine whether the two genes are located on the same operon. Genes located on the same operon often have similar roles. If this were the case for these genes the function of Rv2632c should be taken into consideration as well. RT-PCR suggests that Rv2632c and Rv2633c are not located on the same operon since the intergenic region between the two genes could not be amplified. If the two genes are on separate operons, then this region between genes would be on separate mRNA transcripts. Therefore, when the primers bind to the sequence, they are not on complimentary strands and cannot amplify the fragment. Since microarray data implied otherwise, qRT-PCR was conducted to get a more specific, quantitative measure of this relationship. qRT-PCR suggests that Rv2632c and Rv2633c are both abundant in the cell based on similar relative abundance to the gene *sigA* which is known to be highly expressed in *Mtb*. High expression of Rv2633c is consistent with the
vibrant magenta color of the reporter in *Mtb*. These two genes also appear to be similar in expression levels, which would be consistent with an operon. qRT-PCR will also be performed with RNA from ΔPhoPR and ΔWhiB3 strains. With this RNA, it will be possible to see how the abundance of these genes are affected by mutation in the transcription factors. It is predicted that their relative abundance will drop significantly in the transcription factor mutant if the gene is regulated by it. RNA from *Mtb* during a macrophage infection will also be used for qRT-PCR to determine if the abundance of Rv2633c increases during cell stress.

The qRT-PCR data, the heat map, and the two genes being adjacent suggest they may reside on the same operon. However, the RT-PCR data does not agree with this so further investigation must be conducted to conclude the operon organization of Rv2632c and Rv2633c. Based on the results of qRT-PCR performed with RNA from ΔPhoPR and ΔWhiB3 strains and RNA from *Mtb* macrophage infection, it will be possible to see if the levels of Rv2632c and Rv2633c are coordinately regulated. Confirming operon organization may also include attempting to find a promoter region for Rv2632c or using primer extension to locate the 5’ end of the mRNA transcript.

Similarly, the knock out strain will be used with reverse genetics. Once this construct in transformed into the cell, Rv2633c will not be present any more. By exposing the bacteria to the same environmental stresses described above, we will identify conditions under which Rv2633c is required for the survival and growth of mycobacteria. Based on its sequence homology to hemerythrins and Dps, both proteins involved in stress survival, we hypothesize
that the knock-out mutant will exhibit reduced tolerance for stresses such as hypoxia, acid pH, and starvation.

*Mtb* has many mechanisms still uncharacterized to maintain long term infection. Rv2633c may be one of these mechanisms, using a hemerythrin-like domain to sense hypoxia and oxidative stress coupled to a Dps domain to protect the cell’s DNA during these conditions. By defining novel *Mtbc* virulence mechanisms like Rv2633c, it will be possible to develop new drugs that will target TB in a way to improve the quality of life for any people, and save millions of lives every year.
APPENDIX: PRIMER CATALOG
<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’-3’)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2633pro_mCh_F</strong></td>
<td>TAGAGGATCGTCGGC ACC/CGGGTGGGCAG GAGCT</td>
<td>Forward primer amplifying the Rv2633c promoter region ~1,000 bp upstream of the gene.</td>
</tr>
<tr>
<td><strong>2633pro_mCh_R</strong></td>
<td>TCGCCCTTGCTGACCA T/TTCGTTGCTCTATTC CGTATTCG</td>
<td>Reverse primer amplifying the Rv2633c promoter region ~1,000 bp upstream of the gene.</td>
</tr>
<tr>
<td><strong>2633c/32cRT_F</strong></td>
<td>GACGTAGAGGAACGC GACAT</td>
<td>Forward primer for amplifying the intergenic region between Rv2632c and Rv2633c</td>
</tr>
<tr>
<td><strong>2633c/32cRT_R</strong></td>
<td>CAAACCAACACCCACC ATCT</td>
<td>Reverse primer for amplifying the intergenic region between Rv2632c and Rv2633c</td>
</tr>
<tr>
<td><strong>2633c_Dn_F</strong></td>
<td>CGGTTGTCAGTGCTA GGC</td>
<td>Forward primer for amplifying the region directly downstream of Rv2633c for creating the knock out strain</td>
</tr>
<tr>
<td><strong>2633_Dn_R</strong></td>
<td>CCTCGGTCAGCAGTT GGC</td>
<td>Reverse primer for amplifying the region directly downstream of Rv2633c for creating the knock out strain</td>
</tr>
<tr>
<td><strong>2633_Up_F</strong></td>
<td>GACGTGGCGGGATGT TAATC</td>
<td>Forward primer for amplifying the region directly upstream of Rv2633c for creating the knock out strain</td>
</tr>
<tr>
<td>Primer Name</td>
<td>Sequence</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>2633_Up_R</td>
<td>CGAATACGGAATAGA GCAACGA</td>
<td>Reverse primer for amplifying the region directly downstream of Rv2633c for creating the knock out strain</td>
</tr>
<tr>
<td>pFCKO_dn_F</td>
<td>GTGAGTCGTATTACA ATTCACTGG</td>
<td>Forward primer for opening the pFCKO vector downstream of the Hyg marker</td>
</tr>
<tr>
<td>pFCKO_dn_R</td>
<td>CGCTAAATTCAATTCG CCCTAT</td>
<td>Reverse primer for opening the pFCKO vector downstream of the Hyg marker</td>
</tr>
<tr>
<td>pFCKO_Up_F</td>
<td>GGGACTAGTCCTGCA GGTTTAAA</td>
<td>Forward primer for opening the pFCKO vector upstream of the Hyg marker</td>
</tr>
<tr>
<td>pFCKO_Up_R</td>
<td>CCAAGCTCAGAATTA ACCCTCACTAAA</td>
<td>Reverse primer for opening the pFCKO vector upstream of the Hyg marker</td>
</tr>
<tr>
<td>M13F</td>
<td>CAGGAAACAGCTATG ACCATGA</td>
<td>Forward primer used for screening the knockout for the upstream fragment, hyg marker, and downstream fragment</td>
</tr>
<tr>
<td>M13R</td>
<td>CGTCGTTTTACA</td>
<td>Reverse primer used for screening the knockout for the upstream fragment, hyg marker, and downstream fragment</td>
</tr>
<tr>
<td>pVVmCh_FC_Rpr</td>
<td>GGTGCCGACGATCCT CTA</td>
<td>Reverse primer for amplifying during Round the \textit{Horn} PCR</td>
</tr>
<tr>
<td>SigA_F</td>
<td>CGACGCTGAACCAGA</td>
<td>Forward primer used as a control for qRT-PCR</td>
</tr>
<tr>
<td>Primer Name</td>
<td>Sequence</td>
<td>Function</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>SigA_R</td>
<td>TTCGAGGTCTTCTGTG  GTCTT</td>
<td>Reverse primer used as a control for qRT-PCR</td>
</tr>
<tr>
<td>Rv2632c_F</td>
<td>ATTGATGTGCTGATTG  AGGTTG</td>
<td>Forward primer for amplifying inside the Rv2633c gene during qRT-PCR</td>
</tr>
<tr>
<td>Rv2632c_R</td>
<td>ATATCCGACGACGTCA  AAAGC</td>
<td>Reverse primer for amplifying inside the Rv2633c gene during qRT-PCR</td>
</tr>
<tr>
<td>Rv2633_F</td>
<td>ATGCCTACGACGTATT  AAAGCG</td>
<td>Forward primer for amplifying inside the Rv2632c gene during qRT-PCR</td>
</tr>
<tr>
<td>Rv2633_R</td>
<td>GAACACGTTCCACTCT  TCTTCA</td>
<td>Reverse primer for amplifying inside the Rv2632c gene during qRT-PCR</td>
</tr>
</tbody>
</table>
REFERENCES


