Characterization of Hemerythrin-like Protein Rv2633c

2016

Michelle D. Cherne

University of Central Florida

Find similar works at: https://stars.library.ucf.edu/honorstheses

University of Central Florida Libraries http://library.ucf.edu

Part of the Biochemistry Commons, Bioinformatics Commons, and the Molecular Biology Commons

Recommended Citation

Cherne, Michelle D., "Characterization of Hemerythrin-like Protein Rv2633c" (2016). Honors Undergraduate Theses. 11.
https://stars.library.ucf.edu/honorstheses/11

This Open Access is brought to you for free and open access by the UCF Theses and Dissertations at STARS. It has been accepted for inclusion in Honors Undergraduate Theses by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
CHARACTERIZATION OF HEMERYTHRIN-LIKE PROTEIN Rv2633C

by

MICHELLE CHERNE

A thesis submitted in partial fulfillment of the requirements
for the Honors in the Major Program in Biotechnology
in the College of Medicine and in the Burnett Honors College
at the University of Central Florida

Spring Term, 2016

Thesis Chair: William Self, Ph.D.
Abstract

Hemerythrin-like protein Rv2633c is a small 18 kDa protein that is expressed in *Mycobacterium tuberculosis* (*Mtb*). Sequence analysis of Rv2633c predicts the presence of a hemerythrin-like domain, which binds dioxygen using a µ-oxo-bridge (Fe-O-Fe), rather than a heme group. Though it is noticeably upregulated during macrophage infection and during *in vitro* acidification, the role of Rv2633c in *Mtb* survival has yet to be elucidated. This project aims to characterize the function of Rv2633c by studying the *in vitro* response of the recombinant protein to conditions present in the macrophage lysosome, such as reduced oxygen levels or the presence of reactive oxygen species. UV-visible spectroscopy is used to observe these changes, as the spectrum shows a characteristic peak at 330 nm that likely corresponds to the diiron cofactor in its native state. Our results show this spectrum shifts in response to hydrogen peroxide addition, showing the proposed environmental conditions can affect the active site. Bioinformatics techniques, such as the 3D modeling program SWISS-MODEL, have been used to hypothesize possible structure and function. Determining the function of Rv2633c may help explain how *Mtb* so readily evades the human immune system to reside in the macrophage.
Acknowledgements

I would like to thank Dr. Kyle Rohde, Dr. Victor Davidson, and Dr. Dmitry Kolpashchikov for donating their time to be on my committee and for their advice and expertise. I would also like to thank Steven McKenzie for successfully cloning the expression vectors I was able to use to purify Rv2633c, and Michael Boring, for determining optimum concentrations of IPTG for induction. And thank you to my parents and my friends for always believing in me. I would like to thank my Committee Chair, Dr. William Self, for being an incredible mentor. I couldn’t be more grateful for your constant encouragement and support.
# TABLE OF CONTENTS

Introduction ......................................................................................................................... 1

Rv2633c .................................................................................................................................. 1

*Mycobacterium tuberculosis* ................................................................................................. 1

*Mtb* Survival in the Host ................................................................................................. 2

Hemerythrin ....................................................................................................................... 4

Hemerythrin-like Domains in Biology.................................................................................. 7

Materials and Methods ........................................................................................................ 8

Sequence Analysis ............................................................................................................... 8

Protein Structure Analysis ................................................................................................. 8

Expression Protocols ........................................................................................................... 8

Purification of Rv2633c ....................................................................................................... 9

Protein Quantification ......................................................................................................... 9

Oxygen Binding Experiments ............................................................................................ 10

Nitric Oxide Binding Experiments ..................................................................................... 10

Hydrogen Peroxide Experiments ....................................................................................... 10

Redox State Experiments .................................................................................................. 11

Results ................................................................................................................................... 12

Protein Structure Prediction and Modeling ........................................................................ 12
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence Analysis</td>
<td>12</td>
</tr>
<tr>
<td>Predicted Secondary Structure</td>
<td>19</td>
</tr>
<tr>
<td>3D Structural Modeling</td>
<td>20</td>
</tr>
<tr>
<td>Model Prediction Accuracy</td>
<td>26</td>
</tr>
<tr>
<td>Cofactor Binding Predictions</td>
<td>31</td>
</tr>
<tr>
<td>Site Directed Mutagenesis Predictions</td>
<td>38</td>
</tr>
<tr>
<td>Purification and Expression of Recombinant</td>
<td>38</td>
</tr>
<tr>
<td>Response to Relevant Conditions of the Macrophage</td>
<td>42</td>
</tr>
<tr>
<td>Oxygen Binding</td>
<td>42</td>
</tr>
<tr>
<td>Hydrogen Peroxide Addition and Dialysis</td>
<td>43</td>
</tr>
<tr>
<td>Reducing Agent Addition</td>
<td>47</td>
</tr>
<tr>
<td>Conclusions</td>
<td>50</td>
</tr>
<tr>
<td>Sequence Analysis and Structural Modeling Predictions</td>
<td>50</td>
</tr>
<tr>
<td>Purification and Expression</td>
<td>51</td>
</tr>
<tr>
<td>In vitro Macrophage Conditions</td>
<td>52</td>
</tr>
<tr>
<td>Discussion</td>
<td>53</td>
</tr>
<tr>
<td>Citations</td>
<td>55</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1: Sequence Alignment of Rv2633c with Related Mycobacteria Species .......... 15
Figure 2: Phylogenic tree of full protein sequence ......................................................... 16
Figure 3: Phylogenic Tree of Rv2633c with homologs based on comparisons using amino acid sequence 116-161 .................................................................................................................. 17
Figure 4 : Phylogenic Tree of Rv2633c with homologs based on comparisons using amino acid sequence 138-161 .................................................................................................................. 18
Figure 5: Predicted Secondary Structure ..................................................................... 19
Figure 6: CLUSTAL Multiple Sequence Alignment of Structural Template Sequences. 21
Figure 7: Predicted Structure of Rv2633c C-terminus ................................................. 22
Figure 8: Predicted Model from Template Protein Hypothetical protein NMB1532 ...... 24
Figure 9: Predicted Model from Template Protein Hemerythrin-like Domain Protein DcrH.............................................................................................................................................. 24
Figure 10:Predicted Model from Template Protein Bacteriohemerythrin from Methylococcus capsulatus .............................................................................................................................................. 25
Figure 11: Predicted Model from F-box/LRR-repeat protein 5......................................... 25
Figure 12: Hypothetical protein NMB1532 Error Prediction Model............................... 28
Figure 13: Hypothetical protein NMB1532 Error Prediction by Amino Acid Residue ... 28
Figure 14: Hemerythrin-like domain protein DcrH Error Prediction Model................. 29
Figure 15: Hemerythrin-like domain protein DcrH Error Prediction by Amino Acid Residue.......................................................................................................................................... 29
Figure 16: Bacteriohemerythrin (Methylococcus capsulatus) Error Prediction Model.... 30
Figure 17: Bacteriohemerythin (*Methylococcus capsulatus*) Error Prediction by Amino Acid Residue .......................................................... 30

Figure 18: F-box/LRR-repeat protein 5 Error Model .......................................................... 31

Figure 19: F-box/LRR-repeat protein 5 Error Prediction by Amino Acid Residue ........ 31

Figure 20: Monoazido-µ-oxo-diiron binding model of Rv2633c generated from the structural template hemerythrin-like domain protein DcrH .......................................................... 34

Figure 21: µ-oxo-diiron binding model of Rv2633c generated from the structural template hemerythrin-like domain protein DcrH .......................................................... 35

Figure 22: Predicted Mn²⁺ Binding site of Rv2633c from Template Sequence Hypothetical protein NMB1532 .......................................................... 36

Figure 23: Chloro-diiron-oxo binding site predictions of Rv2633c generated with template model Bacteriohemerythin (*Methylococcus capsulatus*) ........................................ 37

Figure 24: Stained SDS-PAGE of Rv2633c. ................................................................ 39

Figure 25: UV-Visible Absorbance Spectrum of Rv2633c Expression and Purification Methods .................................................................................................... 40

Figure 26: Hydrogen Peroxide Addition Causes Increase in Absorbance at 330 nm and 550 nm .......................................................... 44

Figure 27: Overnight Dialysis of Hydrogen Peroxide Shows Spectral Shifts are Reversible ........................................................................................................ 45

Figure 28: Change in 330 nm Absorbance Increases Rapidly upon Hydrogen Peroxide Addition ........................................................................................................ 46

Figure 29: Change in 550 nm Absorbance Increases Rapidly within 2-4 min after Hydrogen Peroxide Addition ........................................................................ 47
Figure 30: Addition of 10 mM Glutathione causes Light Scattering that cannot be Reversed by Dialysis ..................................................................................................................... 48

Figure 31: Addition of 10 mM TCEP causes Light Scattering and Aggregation that cannot be Reversed by Dialysis ........................................................................................................ 49
LIST OF TABLES

Table 1: Structural Templates ................................................................................................................. 21
Table 2: Generated Structural Models of Rv2633c .............................................................................. 23
Table 3: MODFOLD4 Model Prediction Accuracy ................................................................................ 26
Table 4: COFACTOR Ligand Binding Prediction ..................................................................................... 32
Table 5: Protein Concentrations, Measured Absorbance, and Ratios .................................................. 42
INTRODUCTION

Rv2633c

Rv2633c is a small 18 kDa protein highly expressed by \textit{Mycobacterium tuberculosis} (\textit{Mtb}) when taken into the lysosomal compartments of the human macrophage. Microarray data by Dr. Kyle Rohde shows Rv2633c upregulated during TB infection of the macrophage and during \textit{in vitro} acidification to pH 5.5-6.5 \cite{1}. Dr. Rohde and his team have also noted the possible importance of the regulators WhiB3, PhoPR, and TrcRS to the expression of Rv2633c.

An interesting feature of Rv2633c is a hemerythrin-like domain. This domain is known to bind oxygen using a diiron center. This domain is found in a wide range of organisms and have been shown to have functions such as oxygen binding, iron sequestration, and chemotaxis. Many of the known functions of hemerythrin-like proteins could be relevant to the environmental conditions within the hostile environment of the host macrophage.

\textit{Mycobacterium tuberculosis}

In just the past 200 years, \textit{Mtb} has claimed the lives of around one billion people \cite{2}. Currently, approximately 1/3 of the world’s population is infected with latent \textit{Mtb}, which can cause tuberculosis disease if the infected person’s immune system becomes compromised \cite{3}. The World Health Organization reported that in 2013, 9 million people developed tuberculosis (480,000 of these cases being multidrug resistant TB) while 1.5 million people died from tuberculosis \cite{4}. Though great progress has been made towards treatment of tuberculosis, more progress must still be made in developing more effective treatments to stop the spread of this epidemic.
When *Mtb* enters the lungs through aerosol droplets, macrophages take in the foreign bacteria through phagocytosis [5]. Throughout its coevolution with the human species, *Mtb* has developed ways to survive the hostile environment of the phagolysosome by preventing its maturation [6]. Due to *Mtb*’s complex cell surface structure it can be taken up by multiple macrophage receptors [7]. The receptor that *Mtb* is taken up by can greatly affect the bacteria’s survival, for instance, the uptake of *Mtb* by the CR3 receptor prevents the activation of the macrophage and increases the bacteria’s chance of survival within the macrophage [8]. One method of preventing phagosome lysosome fusion used by *Mtb* is blocking the accumulation of the membrane component phosphatidylinositol 3-phosphate (PI3P) onto the phagosome [9, 10]. This component is thought to bind several proteins involved in the fusion of the phagosome and lysosome [11].

**Mtb Survival in the Host**

*Mtb* can also remain alive within the macrophage phagolysosome even as the pH becomes acidic, usually between pH 5.0 and 4.5 [12]. *Mtb* has also been shown to multiply even when inside the phagolysosome [6]. Mycobacterium species have even been found surviving in the volcanic springs of Yosemite National park at pH as low as 1 [13]. An analysis of the rRNA content of highly acidic endolithic environments of Yellowstone national park showed that 37% of species were Mycobacterium [13]. Some species of bacteria such as *Listeria monocytogenes* even require the acidity of the lysosome for survival or for the activation of virulence factors [14]. There are many possible methods used by *Mtb* to maintain pH homeostasis in acidic conditions. *Mtb* most likely uses some of the common methods used by most bacteria, such as cation pumps to regulate internal pH, or a urease to produce ammonia [15-17]. The tough cell wall of *Mtb* is also a likely defense against acidic environments. The bacteria’s cell wall has a
bilayer membrane incased by a peptidoglycan layer that is covalently bonded to long mycolic acids [18].

*Mtb*, like essentially all other organisms, requires iron for growth, and free iron a limiting nutrient in the host. Most (70-75%) of iron in the human body is bound to the porphyrin, which eventually forms heme [19]. The remaining iron not bound to heme is bound to the proteins transferrin, lactoferrin, or ferritin [20]. *Mycobacterium tuberculosis* is known to sequester iron using compounds called siderophores, which is a type of compound that strongly binds Fe$^{3+}$ [21]. *Mtb* has two main siderophores: mycobactin T and carboxymycobactin [22]. Mycobactin binds iron intracellularly while carboxymycobactin binds iron extracellularly [23]. Carboxymycobactin can remove the iron from both ferritin and transferrin [24]. Rv1436, or glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase, exists on the cell surface of *Mtb* and directly binds human transferrin and internalizes it, without the need for siderophores [25].

All aerobic organisms must have defenses used to scavenge reactive oxygen species, such as superoxide and hydrogen peroxide [26] [27]. The damaging effects of reactive oxygen species have been utilized by many organisms as a defense against pathogens [28]. *Mtb* has many known strategies to combat these reactive oxygen species. The bacteria uses the enzymes catalase and superoxide dismutase, as well as peroxiredoxins to defend against the reactive environment of the host macrophage [29] [30]. A unique method used by mycobacteria is the use of mycothiol, which is a thiol present within the cytoplasm that create a reducing environment that helps defend against hydrogen peroxide [31].

*M. tuberculosis* and other related species of mycobacterium such as *M. leprae* can inhibit the fusing of the phagosome with the lysosome [32], and therefore can remain in an environment more suitable to survival at a more mild pH of around 6.2 [12]. After initial entry into the
phagosome, the phagosome buds off from the membrane and forms smaller vacuoles that cannot fuse with lysosomes, protecting the bacteria [33]. Although much is known about bacterial survival in the macrophage, many questions remain.

The genes *whiB3* and *phoPR*, have been shown to be upregulated in response to low pH, while *trcR* is upregulated during the initial stages of *Mtb* infection in humans [34-36]. *WhiB* is a gene family of seven genes (*whiB1-7*) that express transcription factors under varying conditions of environmental stress to mycobacterium [37]. *WhiB* is expressed 12 fold higher when the cell is introduced to the low pH similar to the phagosome environment [38]. *PhoPR* is a two component regulatory system that is involved in virulence and lipid synthesis for the cell envelope [39]. Strains of *Mtb* that have an inactivated *phoPR* gene show considerably diminished ability to grow within macrophage like cells [39]. The *phoPR* regulatory system also appears to sense pH, as it modulates expression of several environmental pH dependent genes [40]. *TrcR* is a regulator that binds and inhibits Rv1057 in *Mtb*, which is a protein that has a beta-propeller fold structure and is involved in early growth within the human macrophage [41]. Little is known about its structure or function. Preliminary evidence suggests these transcriptional regulators each play a role in the regulation of the gene encoding Rv2633c (Preliminary data, Rohde).

**Hemerythrin**

Hemerythrins are metalloproteins that reversibly bind dioxygen using two iron atoms [42]. Unlike most other dioxygen binding proteins hemerythrin contains no heme cofactor. The molecular weight of the hemerythrin octamer of the worm species *Pectinaria gouldii* is 107,000, and the subunits have molecular weights of 13,500, and most other hemerythrins have similar molecular weights [43]. They are most often found in various types of marine invertebrates. The
hemerythrins of the peanut worm, *Sipunculida*, have been the most extensively studied of this group [44]. The hemerytherin proteins are usually octomeric in the blood and monomeric when in the muscle cells [45]. The monomer is referred to as myohemerythrin. Other oligomers including dimeric, trimeric, and tetrameric has also been found [42]. It is thought the hemerythrins and myohemerythrins play the role in these invertebrates that hemoglobin and myoglobin play in other species [44].

Myohemerythrin has four approximately parallel alpha helices that surround the two iron atoms [46]. The binding site within the channel created by the parallel alpha helices contain hydrophobic amino acid residues that likely serve to create a binding environment for O₂ [42]. The N terminal has a disordered, nonhelical region [46]. The protein connects to the iron atoms at 6 residues, with 3 residues to each iron atom. A residue from each of the 4 alpha helices coordinates in a perpendicular plane with the iron atoms [46]. Access to the diiron center is obstructed by the N-terminal arm and a side chain of one of the helices [46].

The diiron core in the active site of hemerythrin proteins from marine invertebrates are connected by a \( \mu \)-oxo-bridge (Fe-O-Fe) and two carboxylate groups from corresponding aspartate and glutamate residues [47]. Three histidine residues coordinate with Fe1 and two histidine residues and a glutamate residue coordinate Fe2 [48]. The diiron center can usually be found in the deoxy, met, or oxy state. The \( \mu \)-oxo bridge is retained in each oxidation state [49]. When in the deoxy form, the oxidation states are Fe (II, II), but they transition to Fe (III, III) in the oxy state [50]. In the deoxy state the bridging oxygen is protonated [49]. The proposed mechanism for O₂ binding is the proton from the bridging oxygen is transferred to O₂ to stabilize bound to Fe in the peroxy state [49]. When the Fe-O-Fe center is not bound to oxygen in the deoxy state, the Fe-O distances are 1.8 Å and the Fe-Fe distance is 3.57 Å. When O₂ is bound the
Fe-O distances are 2.0 Å and the Fe-Fe distance is 3.24 Å [49]. The dioxygen is bound in a hydroperoxy state, rather than a dioxygen complex which is quite different from usual dioxygen binding molecules. Hemerythrin can also bind to other ligands similar to O₂ such as nitric oxide, or azide [42].

Hemerythrins were first found in certain species of marine invertebrates; Sipuncula (peanut worm), Priapulida, Brachiopoda, and several different annelids [51, 52]. Similar proteins have been found scattered throughout bacterial species and have even been found in archaea. This leads some to believe that hemerythrin may be a protein that was acquired horizontally in bacteria and archaea [44].

An evolutionary biology study published by Xavier Bailly and his team in 2008 searched for the hemerythrin domain throughout hundreds of different genomes within eukaryotes, archaea, and, bacteria. Along with the already known marine invertebrates, 13 species of fungi were found to contain hemerythrin, though they differed substantially in sequence when compared to the original motif [53]. 43 archaeal species were studied, and only 4 species contained the hemerythrin motif [53]. 444 bacterial genomes were analyzed, and 118 different species contained the hemerythrin motif [53]. 242 of the proteins were single domain and 84 were domains on chimeric proteins [53]. Of these chimeric proteins, 32 were domains of chemotaxis proteins. It should be noted that oxygen requirements of the bacterial species had no correlation to the presence of hemerythrin [53].

Many of the sequences found to align with hemerythrin were altered slightly at important residues. The most common differences were changes in one of the 4 alpha helices or a substitution in one of the five histidine residues which coordinate the two iron atoms. The most common of these substitutions was a glutamine replacing the histidine [53].
The hemerythrin protein motif was shown to be sporadically present throughout all domains of life, but is obviously overshadowed by hemoglobin, especially in highly evolved, multicellular organisms. The same study suggests that hemerythrin may have lost evolutionary favor due to its greater potential for damaging mutations, as it needs seven different residues placed correctly in order to bind the iron atoms.

Hemerythrin-like Domains in Biology

It has become apparent that hemerythrin-like domains have been adapted for a broad range of functions in an even broader range of organisms. The first hemerythrin protein studied in bacteria was McHr, a hemerythrin protein from Methylococcus capsulatus [54]. It is predicted as an O₂ transporter to deliver oxygen to particulate methane monooxygenase (pMMO) for methane oxidation [54]. Desulfovibrio vulgaris is an anaerobic bacterium uses a hemerythrin-like domain as a chemotaxis protein. The hemerythrin-like domain senses the presence of O₂, which triggers a cascade that alters the swimming behavior of the cell away from oxygen [55]. The ovohemerythrin protein YP14 is hypothesized to serve as an iron storage protein during the development of Theromyzon tessulatum, a species of leech [56]. A hemerythrin-like protein found in Mycobacterium smegmatis, msmHr, has recently been shown to regulate the expression of the sigma factor SigF in response to hydrogen peroxide rich environments. It appears to downregulate the production of SigF when hydrogen peroxide is not present, as over expressed msmHr downregulated the expression of SigF and caused increased susceptibility to oxidative stress [57].
MATERIALS AND METHODS

Sequence Analysis

NCBI Basic Local Alignment Search Tool (BLAST) sequencing was used to compare the hemerytherin-like domain in Rv2633c to conserved sequences in order to predict its structure and function [58]. NCBI COBALT was used to create a multiple sequence alignment of proteins with sequences most related to Rv2633c. ExPASy ProtParam peptide properties calculator was used to analyze the amino acid content of the wild type protein.

Protein Structure Analysis

The secondary structure prediction program CFSSP (Chou & Fasman Secondary Structure Prediction Server) was used to analyze the possible secondary structure of Rv2633c [59, 60]. The SWISS-MODEL program was used to determine a probable structure of the protein and active site based on similarly conserved protein structures [61, 62]. The program MODFold 4 was used to independently access the probability of the constructed models fitting the native structure of the protein [63-65]. The program COFACTOR was used to determine the possible ligand binding sites of the generated protein models [66-68]. These results were used to determine possible mutagenesis sites to target for further studies.

Expression Protocols

Rv2633c was expressed in BL21(DE3) E. coli using the well described T7 promoter expression system [69]. The expression time and temperature varied, and was either overnight at room temperature or 3 hours at 37°C. The pET100 expression vector used for each purification produces a tagged protein that includes an N-terminal histidine tag for protein purification. The gene is only expressed in E. coli strains that contain the T7 polymerase gene, such as derivatives
of strain BL21(DE3), under control of the lac promoter/operator region the lambda attachment
region. Large scale expression was undertaken after small scale expressions of around 100ml to
1 L showed sufficient soluble protein for purification. The cells were lysed by French press in 50
mM tris pH 8.0 with 10 mM lysis solution buffer and the supernatant was obtained by
ultracentrifugation for 10 minutes at 40,000 x g.

**Purification of Rv2633c**

A cobalt affinity column was used for affinity purification. The column (approximately
10 ml) was run using a low pressure peristaltic pump and using a 50 mM Tris buffer (pH 8.0).
After initial application of the sample to the column, it was washed with 10 column volumes of
buffer. 25 mM imidazole (in tris buffer) in 50 mM tris buffer pH 8.0 was used to elute the
recombinant protein from the column. In separate purification procedures, 25 mM EDTA was
added to the column, rather than imidazole, to strip the cobalt from the column, and therefore the
bound recombinant protein, from the column. This was done to avoid using imidazole, which
evidence suggests may strip the cofactor out of the protein. Following elution, the protein was
diaalyzed with 3.5 kDa dialysis tubing overnight at 4°C in 3 liters 50 mM Tris buffer pH 8.0, to
remove the imidazole and EDTA, respectively. The purified protein from each expression and
purification was aliquoted into 0.5 ml samples and frozen at -80°C.

**Protein Quantification**

SDS-PAGE was used as a qualitative method to determine the purity of Rv2633c. The
purified protein was quantified using the guanidine hydrochloride method. This was done with 6
M guanidine HCl in 0.05 M tris pH 8.0. The correlated extinction coefficient will be determined
by Protparam, a protein sequence analysis online tool. The absorbance at 280 nm was measured
using UV spectrophotometry to determine the concentration of various protein volumes.
Quantitative Bradford assay was also used to determine the relative amount of protein for each expression and purification of Rv2633c.

**Oxygen Binding Experiments**

A one ml anaerobic quartz cuvette containing the purified Rv2633c was used to study the protein’s response to anaerobic conditions. Nitrogen was sparged into the cuvette for 30 minutes. Oxygen was added back into the anaerobic protein in 50 µl intervals to determine the protein’s ability to bind oxygen. Spectrophotometry was used to observe possible spectral changes between anaerobic and aerobic conditions.

**Nitric Oxide Binding Experiments**

In order to further study the possible binding of Rv2633c, SNAP (S-Nitroso-N-acetylpenicillamine) was added to produce nitric oxide, which often binds to metal containing cofactors also known to bind oxygen. 0.1 mM SNAP in 90% ethyl alcohol was added to 1 ml of recombinant Rv2633c. After 10 minutes of no noticeable activity the molarity of SNAP was increased to 1 mM. UV-visible absorbance spectrometry was used to measure changes in absorbance of the protein.

**Hydrogen Peroxide Experiments**

Hydrogen peroxide was added to the protein in varying concentrations and changes in spectrum were measured with UV visible spectrophotometry. Various concentrations of hydrogen peroxide were used to determine the true reactivity of Rv2633c with reactive oxygen species.

To observe the reaction of Rv2633c under high concentrations of hydrogen peroxide, 50 µl of 3% hydrogen peroxide was added to 950 µl of 3.2 mg/ml Rv2633c (concentration
determined by guanidine denaturation). Measurements of UV-visible absorbance were taken at minute scale time points up to 45 minutes to observe the complete changes in spectrum. BSA as well as water were used to repeat the experiment with the same volumes and concentrations to serve as negative controls. After the reaction came to completion, Rv2633c was dialyzed in 3.5 kDa dialysis tubing in 1.5 liters of 50 mM tris pH 8.0 overnight at 4º and then reexamined by UV-visible spectrometry to test for reversibility of the reaction.

To measure kinetic changes in absorbance produced by hydrogen peroxide addition, the same concentrations and volumes of hydrogen peroxide and protein were used as above. UV-visible absorbance spectrum measurements were taken every 15 seconds for 3 minutes. The experiment was repeated with matching concentrations using water and BSA as controls.

**Redox State Experiments**

The reducing agents glutathione and TCEP (tris(2-carboxyethyl)phosphine) were used to study the effects of reducing conditions on Rv2633c. 0.1 mM of both TCEP and glutathione was added to 1 ml of protein, and increased to 1 mM and then 10 mM after 10 minutes with no noticeable reaction. The UV-visible absorbance spectrum was measured in order to measure the response of the protein to the reducing agents. After the addition of 10 mM reducing agent, the protein was dialyzed with 3.5 kDa dialysis tubing overnight at 4ºC in 50 mM tris buffer pH 8.0, in order to determine whether or not results were reversible upon removal of the reducing agents.
RESULTS

Protein Structure Prediction and Modeling

Sequence Analysis

NCBI BLAST Protein alignments of Rv2633c excluding *Mtb* returns mainly highly similar sequences of hypothetical proteins in other closely related mycobacterium species. NCBI BLAST protein search excluding all mycobacterium species predicts the most similar sequence identity as 44%. The majority of similar sequences are hypothetical proteins with hemerythrin and hemerythrin-like domains, with no known structure or function. Sequence analysis of Rv2633c clearly demonstrates the unique nature of this protein. While the unique nature of Rv2633c makes it an exciting candidate for further study, it also creates challenges in understanding of the protein structure and its function in *Mtb*.

NCBI BLAST of the full sequence of Rv2633c matches with high probability for a hemerythrin HHE cation binding domain, as well as the hemerythrin-like superfamily. This domain is identified for residues 2-116 and 5-115, respectively. However, other interesting information was obtained by analysis of smaller portions of the full sequence. BLAST sequence analysis followed by generation of a phylogenetic tree based on the multiple alignment shows a homology match with the ferritin-like superfamily domain. This protein superfamily is characterized by the ability to catalyze dioxygen-dependent oxidation-hydroxylation using a diiron center [70]. However, the ferritin like superfamily has a 4 alpha helix protein structure, and the 45 amino acid sequence matched is likely only a portion of this domain.

COBALT was used to generate a protein sequence alignment of related species of mycobacteria using the full sequence of Rv2633c, as well as smaller portions located at the C-
terminus that seemed more isolated to mycobacterium. This program was also used to analyze the sequence similarities of the proteins used as templates for structural predictions of Rv2633c.

Analysis of related mycobacterium species by sequence similarity and predicted phylogeny showed that sequence similarity correlated with the degree of pathogenicity of the mycobacterium species. Mycobacterium species that are well established pathogens were more closely related than those that are considered to be opportunistic pathogens in the host (figure 2). The full protein sequence was analyzed as well as two different lengths of the C-terminus, as this region of the protein sequence is more unique to mycobacterium. BLAST sequence analysis returns only one mycobacterium species known to be nonpathogenic as a significant alignment, *Mycobacterium indicus pranii* (figure 2). This nonpathogenic species of mycobacterium had the least sequence similarity when compared to Rv2633c from *Mycobacterium tuberculosis* at a sequence identity of 56%. BLAST sequencing and phylogeny using COBALT alignment shows only opportunistic and fully pathogenic mycobacteria have sequence homology to the 45 amino acids at the carboxyl terminus that do not correspond to the predicted hemerythrin domain, the region in which a ferritin-like domain exists (figure 3). The same process was repeated with just the last 23 amino acids on the carboxyl terminus, as portion of the sequence could not be matched with any structurally homologous models when generating hypothetical 3D structural models (figure 4). This small portion of the sequence further narrows down the range of related mycobacterium species. A small subset of opportunistic and fully pathogenic mycobacterium species were shown as homologues to this portion of Rv2633c. The figures below show a sequence alignment of all conserved mycobacterium species generated from the NCBI databases, each phylogenetic tree was generated from this alignment. Phylogenetic trees generated from the
blast results show the evolutionary distances predicted from mycobacterium species with conserved protein sequences.
Figure 1: Sequence Alignment of Rv2633c with Related Mycobacteria Species
Figure 2: Phylogenetic tree of full protein sequence
Figure 3: Phylogenic Tree of Rv2633c with homologs based on comparisons using amino acid sequence 116-161
Figure 4: Phylogenetic Tree of Rv2633c with homologs based on comparisons using amino acid sequence 138-161
**Predicted Secondary Structure**

CFSSP generated a prediction of the secondary structure of Rv2633c. Figure 5 shows the most likely secondary structure of the protein sequence [59, 60]. The table shows the amino acid residues by sequence, with H, E, T, and C corresponding to the secondary structures helix, beta strand, turn, and coil, respectively [59, 60]. The most prevalent secondary structure predicted from this program was alpha helical. Figure 5 shows the presence of alpha helices scattered throughout, with the largest between residues 125 and 140. Some small areas predict the presence of beta sheets. Coils and turns are also predicted to be 9.9% of the protein residue structure. These areas likely correspond to areas with more flexible structure.

**Figure 5: Predicted Secondary Structure**
3D Structural Modeling

SWISS-MODEL was used to create probable 3D structures of Rv2633c based on preexisting structural models of homologous proteins taken from the Protein Data Bank. Four models were produced using the homologous protein structures as templates. The selected models were Hypothetical protein NMB1532, F-box/LRR-repeat protein 5, hemerythrin-like domain protein DcrH, and bacteriohemerythrin from *Methylococcus capsulatus*. These selections were made based on predictions made by the SWISS-MODEL program on the degree of structural and sequence similarity. Figure 6 shows an alignment of the four selected protein templates compared to Rv2633c, in the order Rv2633c, NMB1532, Hemerythrin-like domain protein, FBXL5, and *Methylococcus capsulatus* bacteriohemerythrin. The sequences are highly varied, with the only similar alignments being related to the hemerythrin domain portion. The table 1 shows relevant features of the matched templates. These four protein templates were selected for further study based on their sequence identity, template coverage of Rv2633c, and the presence of relevant ligands. The alignment of each structural template used to generate 3D structural predictions shows how different each sequence, the only similarity is the hemerythrin-like domain.
Figure 6: CLUSTAL Multiple Sequence Alignment of Structural Template Sequences

Table 1: Structural Templates

<table>
<thead>
<tr>
<th>Template Name</th>
<th>Sequence Similarity</th>
<th>% Sequence Identity</th>
<th>Ligands</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothetical protein NMB1532</td>
<td>0.30</td>
<td>18.38</td>
<td>2 Mn$^{2+}$</td>
<td>Neisseria meningitidis</td>
</tr>
<tr>
<td>hemerythrin-like domain protein DcrH</td>
<td>0.27</td>
<td>15.93</td>
<td>μ-oxo-diiron</td>
<td>Desulfovibrio vulgaris</td>
</tr>
<tr>
<td>Bacteriohemerythin Leu144F mutant from Methylococcus capsulatus</td>
<td>0.28</td>
<td>15.32</td>
<td>Fe$^{2+}$</td>
<td>Methylococcus capsulatus</td>
</tr>
<tr>
<td>F-box/LRR-repeat protein 5</td>
<td>0.30</td>
<td>20</td>
<td>μ-oxo-diiron</td>
<td>Human</td>
</tr>
</tbody>
</table>

The model generated from Hypothetical protein NMB1532 was rated the most accurate by the SWISS-MODEL program, and also had the greatest amount of amino acid residue
coverage when modeled from the template to Rv2633c. This template was able to model residues 1-138 of Rv2633c.

Though the templates used were rated to be the most similar to Rv2633c, they still had relatively low homology, showing once again the unique sequence of Rv2633c. SWISS-MODEL estimates that a sequence similarity of .5 or over generates a structural model highly similar to the native form. No template was available to model the last 22 amino acid residues on the carboxyl end. This unique portion of this structure may serve a function more unique to *Mtb*. Secondary structure predictions of this unique carboxyl portion suggest an unstructured region, possibly a beta strand, followed by a small alpha helix. Structural modeling of this portion was done using SWISS-MODEL with a protein template from pyruvoyl-dependent arginine decarboxylase. This smaller independent molecule extends the modeled section by 11 amino acid residues and also predicts a beta strand. The 11 amino acid residues remaining on the carboxyl end could not be matched to any known structural template, but are predicted by secondary structure sequence analysis to be an alpha helix.

*Figure 7: Predicted Structure of Rv2633c C-terminus*

No model generated from this program could conserve a metal cofactor binding site or ligand binding site of any kind. Though each model template selected had a bound ligand, the
binding sites were not conserved well enough to be added to the models. This has interesting implications of the active site of Rv2633c, as sequence analysis as well as experimental evidence suggests the existence of a metal cofactor.

The table below shows relevant statistical data about each generated structural model. The GMQE score (Global Model Quality Estimation) estimates the accuracy of the model built with the specified template, and is expressed as a number between 0 and 1, with higher numbers indicating a higher accuracy. The QMEAN4 score is a linear combination of the four QMEAN ranked structural descriptors of a model: solvation, torsion, Cβ distance dependence, and all atom distance dependence. The output is that of a Z-score, with a Z score of zero comparing to a high resolution X-ray model generated from a crystal structure [71].

<table>
<thead>
<tr>
<th>Template Name</th>
<th>GMQE Score</th>
<th>QMEAN4 Score</th>
<th>Amino Acid Range Covered</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothetical protein NMB1532</td>
<td>0.50</td>
<td>-5.96</td>
<td>1-138</td>
<td>0.84</td>
</tr>
<tr>
<td>hemerythrin-like domain protein DcrH</td>
<td>0.40</td>
<td>-4.23</td>
<td>4-119</td>
<td>0.70</td>
</tr>
<tr>
<td>Bacteriohemerythin</td>
<td>0.39</td>
<td>-3.74</td>
<td>4-117</td>
<td>0.69</td>
</tr>
<tr>
<td>F-box/LRR-repeat protein 5</td>
<td>0.45</td>
<td>-7.25</td>
<td>4-136</td>
<td>0.81</td>
</tr>
</tbody>
</table>

The four probable models are shown below, along with a summary of each models’ details. The cartoon models show the structural features of each protein, with blue indicating the N-terminus and red indicating the C-terminus.
Figure 8: Predicted Model from Template Protein Hypothetical protein NMB1532

Figure 9: Predicted Model from Template Protein Hemerythrin-like Domain Protein DcrH
Figure 10: Predicted Model from Template Protein Bacteriohemerythrin from *Methylococcus capsulatus*

![Figure 10](image10)

Figure 11: Predicted Model from F-box/LRR-repeat protein 5

![Figure 11](image11)
**Model Prediction Accuracy**

A separate program, MODFold 4, was used to analyze the probability of the model produced [63-65]. The program ranks structural models against the known model and determines the probability the predicted model is correct. The Global Model Quality Score predicts the likelihood of an accurate model [63-65]. It is a number between 0 and 1, with scores below 0.2 indicating relatively improperly folded models and scores above 0.4 indicating more confident model predictions [63-65]. Each of the four generated models are shown in cartoon figures colored by RasMol temperature coloring, with blue indicating regions of low position uncertainty and areas of red with high position uncertainty. The graphs shown predict the error in Angstroms of the predicted alpha carbon position by the model with the native form of the corresponding of the native structure. This shows regions of the predicted model that likely contain the most error. The green lines show positions where the protein template was not matched with the target sequence [63-65].

**Table 3: MODFOLD4 Model Prediction Accuracy**

<table>
<thead>
<tr>
<th>Model Template</th>
<th>Global Model Quality Score</th>
<th>P-Value</th>
<th>Confidence Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothetical protein NMB1532</td>
<td>0.4461</td>
<td>0.00873</td>
<td>High</td>
</tr>
<tr>
<td>Hemerythrin-like domain protein DcrH</td>
<td>0.3650</td>
<td>0.02029</td>
<td>Medium</td>
</tr>
<tr>
<td>Bacteriohemerythrin <em>(Methylococcus capsulatus)</em></td>
<td>0.3561</td>
<td>0.02225</td>
<td>Medium</td>
</tr>
<tr>
<td>F-box/LRR-repeat protein 5</td>
<td>0.2933</td>
<td>0.02933</td>
<td>Medium</td>
</tr>
</tbody>
</table>
The model produced from NMB1532 was ranked with a global model quality score of 0.4461. The NMB1532 model had the highest probability of being correct, rated with the P-value of 0.00873. Interestingly, NMB152 binds Mn$^{2+}$, rather than diiron derivatives. While this seems counterintuitive due to the hemerythrin domain present in the sequence, the binding site of the manganese cation was not conserved in the generated model. It must also be taken into consideration that the sequence similarity of this protein to Rv2633c is only 18.38%.

Out of the four models, the model produced from Hypothetical Protein NMB1532 was selected as the most probable model. This model also had the highest quality rating predicted by the SWISS-MODEL program. This model most likely shares the greatest overall structural homology.

Although programs exist that can align multiple templates, this technique was not used, in favor of selecting the most probable model. A study on using multiple templates to improve quality of homology models in automated homology modeling showed that using multiple templates on average does little to improve the generated model [72]. Using multiple templates generally only improves models by extending the length of the protein model, but the selected model has the most protein sequence modeled out of the four analyzed.
Figure 12: Hypothetical protein NMB1532 Error Prediction Model

Figure 13: Hypothetical protein NMB1532 Error Prediction by Amino Acid Residue
Figure 14: Hemerythrin-like domain protein DcrH Error Prediction Model

Figure 15: Hemerythrin-like domain protein DcrH Error Prediction by Amino Acid Residue
Figure 16: Bacteriohemerythin (*Methylococcus capsulatus*) Error Prediction Model

Figure 17: Bacteriohemerythin (*Methylococcus capsulatus*) Error Prediction by Amino Acid Residue
Cofactor Binding Predictions

Though a bound ligand was not conserved using the possible templates generated by the SWISSMODEL program, experimental evidence suggests the presence of a metal cofactor. The
The program COFACTOR was used to predict the presence of cofactors, by using the generated 3D structures of Rv2633c and comparing them to related protein cofactor binding sites from resolved structures [66-68]. For this analysis, the generated models from each selected template were matched with ligand binding sites from related protein structures, and the accuracy of the binding site was accessed. The $C_{score}^{LB}$ is the confidence score of the predicted binding site and is a number between 0 and 1, with 1 being the most probable model. The BS score is the sequence and structure similarity between the binding site of the template and the predicted ligand binding site of the model, with a score of greater than 1 indicating a significant match between the two [66-68].

Table 4: COFACTOR Ligand Binding Prediction

<table>
<thead>
<tr>
<th>Model Template</th>
<th>Predicted Ligand</th>
<th>Binding Site Template</th>
<th>$C_{score}^{LB}$</th>
<th>BS Score</th>
<th>Predicted Binding Site Residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemerythrin-like domain protein DcrH</td>
<td>FeA</td>
<td>met-azido-DcrH-Hr</td>
<td>0.61</td>
<td>1.19</td>
<td>11,14,45,46,49,67,71,101,102,105,106,110</td>
</tr>
<tr>
<td>Bacteriohemerythrin (Methylococcus capsulatus)</td>
<td>CFO</td>
<td>Met DcrH-Hr</td>
<td>0.55</td>
<td>1.22</td>
<td>11,45,47,49,67,71,102,105,106,110</td>
</tr>
<tr>
<td>Hemerythrin-like domain protein DcrH</td>
<td>FeO</td>
<td>Deoxy-DcrH-Hr</td>
<td>0.31</td>
<td>1.21</td>
<td>11,45,49,67,71,105,110</td>
</tr>
<tr>
<td>Hypothetical protein NMB1532</td>
<td>MN</td>
<td>Hypothetical protein NMB1532</td>
<td>0.27</td>
<td>1.29</td>
<td>11,45,49,109, 114</td>
</tr>
</tbody>
</table>

The program Swiss-PdbViewer 4.1.0 was used to generate images of the protein models with the predicted bound cofactors. The PDB file was generated from the COFACTOR server. However, the models constructed are high in error, and are used as a reference only, as the model structures do no truly fit the structural requirements to bind the cofactors shown effectively by
the SWISS-MODEL program. Images were generated showing the probable location of the bound cofactor in several different displays. All of the four previously discussed models were tested against the COFACTOR program, and the three most probable ligand binding sites were chosen for further study. Predicted binding sites below 0.25 were not considered, as these likely have low homology to the true native structure of the protein.

The highest scoring binding site similarity was generated from the Hemerythrin-like domain protein DcrH model, which predicted azide bound to a μ-oxo diiron bridge. The second most probable binding site similarity was also generated from the same predicted structural model, but bound with the oxidized form rather than azide. However, it must be taken into account that the structural model generated from DcrH was not the highest predicted similarity model. The third possible cofactor was a Mn$^{2+}$ atom. Though given a confidence score of only 0.27, the model the data is generated from is the highest ranked structural model of Rv2633c. The model generated from F-box/LRR-repeat protein 5 could not be matched to the corresponding ligand known to bind the target protein, showing a low chance of a similar binding site in Rv2633c.

The cartoon image of the protein bound to the cofactor generated by the COFACTOR program is shown, followed by the model with only predicted cofactor binding residues shown.
Figure 20: Monoazido-µ-oxo-diiron binding model of Rv2633c generated from the structural template hemerythrin-like domain protein DcrH
Figure 21: μ-oxo-diiron binding model of Rv2633c generated from the structural template hemerythrin-like domain protein DcrH
Figure 22: Predicted Mn$^{2+}$ Binding site of Rv2633c from Template Sequence Hypothetical protein NMB1532
Figure 23: Chloro-diiron-oxo binding site predictions of Rv2633c generated with template model Bacteriohemerythin (Methylococcus capsulatus)
Site Directed Mutagenesis Predictions

Though the focus of this work was on the wild type protein, the extensive sequencing and predictive structural analysis of wild type Rv2633c showed residues whose mutation may be telling of protein function. Histidine residues 11, 45, and 49 are the immediate choices, as these residues were conserved as predicted binding sites in every significant prediction by COFACTOR. Histidine 105 is also a conserved binding residue in three of the four most probable binding site predictions. Histidine residues are also known to be the main amino acid associated with the μ-oxo diiron binding site of hemerythrin proteins [42]. Substitution of these histidine residues would render the protein unable to bind the predicted cofactor. Alanine 106 may also serve as a useful target for mutagenesis, as it seems to play some role in shaping the binding pocket of the predicted cofactor. Leucine 101 may also have a similar function. Histidine residue or glutamate residues 126, 128, or 129 could test the possibility of the unique ‘ferritin-like superfamily’ c terminus of this protein playing some role in its function, as ferritin-like proteins are known to bind diiron, or in some cases manganese, using glutamate and histidine residues in close proximity [70].

Purification and Expression of Recombinant

Various expression methods were used to determine optimum conditions for isolation of large amounts of protein, as well as protein with the putative diiron cofactor. Table 5 summarizes the purifications and expressions performed. In each case relatively high concentrations of recombinant protein were obtained. The recombinant protein was highly pure when analyzed qualitatively with SDS-PAGE. SDS-PAGE of each step in the purification also shows a substantial amount of recombinant protein was present in the insoluble portion of the lysate. Whether this is due to unfolded inclusion bodies or an inherent property of the native protein is
unclear. Protein concentration of each preparation was performed by a quantitative Bradford assay in order to compare the relative amount of protein purified from each expression.

Figure 24: Stained SDS-PAGE of Rv2633c.

Starting from left the wells are: Rv2633c, Rv2633c, resolubilized pellet, buffer wash volume 3, buffer wash volume 2, flow through 1, flow through 2, flow through 3, prestained protein ladder (MWs: 170, 130, 95, 72, 55, 43, 34, 26, 17, 10 kDa)
Figure 25: UV-Visible Absorbance Spectrum of Rv2633c Expression and Purification Methods

Hemerythrin proteins characteristically have absorbance in the 330-350 nm range, as well as in the 550 nm, range when in the oxidized form, while has no characteristic absorbance when lacking iron cofactor or in the deoxy state (Fe(II,II)). [73]. Though the expression and purification methods used returned high concentrations of protein each time, the unique absorbance present at 330 nm that correlates to possible presence of cofactor was not
consistently reproduced (Figure 25). The ratio of the absorbance at 280 nm was compared with the absorbance at 330 nm (Table 5). This comparison was made to determine the amount of protein that may contain cofactor to the overall protein purified. By comparing this ratio from each preparation of Rv2633c, it could be determined which purification and expression method was most successful in producing protein with hemerythrin-like absorbance. For a baseline, 700 nm absorbance was chosen to remove any light scattering that may contribute to shorter wavelength measurements.

A lower ratio of 280 nm absorbance to 330 nm absorbance would be ideal, as this likely means more of the cofactor is present in the protein. The 700 nm baseline showed that the absorbance measurements are consistent enough to allow for comparison. The absorbance at 330 nm was compared to the concentration in order to determine if the cofactor absorbance spectrum could be correlated to the amount of protein. The results show that the absorbance at 330 nm, and likely the cofactor bound to Rv2633c, is highly variant in each expression and purification, even when conditions were repeated.

The absorbance at 330 nm depends on variables more complex than expression temperature, molarity of IPTG during induction and purification technique. However, the data shows that purification with imidazole as the elution agent contributed to destabilizing cofactor binding. Iron atoms within a hemerythrin domain are bound by interactions with histidine residues, so it is logical that addition of high concentrations of imidazole could destabilize a bound cofactor. Acidic conditions used for purification also destabilized the protein, this preparation had the highest 280nm/330nm ratio (Table 5). Hemerythrin-like proteins have also been shown experimentally to destabilize in acidic conditions [74].

41
### Table 5: Protein Concentrations, Measured Absorbance, and Ratios

<table>
<thead>
<tr>
<th>Prep</th>
<th>280 nm</th>
<th>350 nm</th>
<th>700 nm</th>
<th>Concentration mg/ml</th>
<th>280/350</th>
<th>330/Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 mM Imidazole Expressed at Room Temperature Overnight</td>
<td>1.25</td>
<td>0.84</td>
<td>0.13</td>
<td>1.129</td>
<td>1.48</td>
<td>0.75</td>
</tr>
<tr>
<td>25 mM Imidazole Expressed at Room Temperature Overnight with 1 mM Iron during growth</td>
<td>0.70</td>
<td>0.19</td>
<td>0.0059</td>
<td>0.844</td>
<td>3.65</td>
<td>0.23</td>
</tr>
<tr>
<td>25 mM Imidazole Expressed at Room Temperature Overnight</td>
<td>0.24</td>
<td>0.12</td>
<td>0.055</td>
<td>0.0280</td>
<td>1.94</td>
<td>4.46</td>
</tr>
<tr>
<td>25 mM Imidazole Expressed at 37ºC for 3 hours</td>
<td>0.51</td>
<td>0.16</td>
<td>0.050</td>
<td>3.973</td>
<td>3.04</td>
<td>0.051</td>
</tr>
<tr>
<td>Acid, MES pH 5, followed by EDTA 25 mM 01.03.16</td>
<td>0.65</td>
<td>0.17</td>
<td>0.049</td>
<td>0.7575</td>
<td>3.94</td>
<td>0.22</td>
</tr>
<tr>
<td>25 mM EDTA Expressed at Room Temperature Overnight</td>
<td>1.29</td>
<td>0.44</td>
<td>0.072</td>
<td>2.313</td>
<td>2.92</td>
<td>0.19</td>
</tr>
</tbody>
</table>

### Response to Relevant Conditions of the Macrophage

**Oxygen Binding**

Hemerythrin like proteins often have the ability to reversibly bind O₂ with a noticeable change in UV-visible absorbance when bound and unbound [42]. If Rv2633c bound O₂ reversibly, anaerobic conditions may have show spectral shifts. Nitrogen purging of the sample in an anaerobic quartz cuvette for 40 minutes showed no noticeable change in the UV-visible absorbance spectrum of Rv2633c. If the 330 nm peak was caused by the presence of reversibly
bound dioxygen there would likely have been some change after the removal of oxygen, as 40 minutes should be sufficient to remove almost all traces of dioxygen from the protein sample.

When SNAP (S-Nitroso-N-acetylpenicillamine) was added to the protein samples to release nitric oxide, no change in the UV-visible absorbance spectrum was observed. Changes in absorbance that did occur were due to the baseline absorbance of SNAP reagent. As dioxygen binding proteins generally bind nitric oxide as well, this result also suggests Rv2633c does not reversibly bind dioxygen.

Hydrogen Peroxide Addition and Dialysis

Hydrogen Peroxide was added in order to study the protein’s response to reactive oxygen species, which are present in the macrophage lysosome. The amount of hydrogen peroxide added was far above conditions that could be found in a biological system, at approximately 0.4 M. However, the results have interesting implications for the cofactor within Rv2633c. H₂O₂ causes a large, reversible UV-visible spectrum shift, as well as changes in the protein color clearly visible by eye. Concentrated Rv2633c containing the cofactor has a pink to brown color. The spectrum shifts to peak absorbance at approximately 550 nm (Figure 26), which is consistent with the absorbance spectrum of a hemerythrin protein in the oxidized form [73]. Controls with BSA and the same concentration of H₂O₂ show some spectral shifts that are barely present, and are likely associated with the presence of H₂O₂.
Figure 26: Hydrogen Peroxide Addition Causes Increase in Absorbance at 330 nm and 550 nm

Upon dialysis, the Rv2633c almost completely regains its original spectrum and color, including the 330 nm peak. This suggests that the changes caused by H₂O₂ are happening to the cofactor, rather than just the result of degradation of the protein.
Measurements of the absorbance spectrum on the seconds scale shows interesting spectral shifts that may be telling of the state of the cofactor. The absorbance at 330 nm rapidly increases upon addition of hydrogen peroxide and then slows after about a minute. The peak at 550 nm begins to appear immediately after the hydrogen peroxide addition but then increases rapidly after about 2 minutes.
Figure 28: Change in 330 nm Absorbance Increases Rapidly upon Hydrogen Peroxide Addition

Rv2633c Absorbance at 330 nm over Time

Absorbance at 330 nm

Time (Seconds)
Upon addition of the same concentration of tert-butyl hydrogen peroxide, there is no change in absorbance spectrum. This suggests the reaction occurring depends on the reactive molecule from hydrogen peroxide must have access to the binding site of the protein in order to occur. Tert-butyl hydrogen peroxide has the same properties as hydrogen peroxide, but with a bulky carbon side group. This is further evidence to suggest the presence of a metal cofactor.

Reducing Agent Addition

The reducing agents TCEP or glutathione were added to Rv2633c to study the effects of reducing equivalents on the protein. No changes in UV-visible spectra were observed up to 1
mM. However, at 10 mM, light scattering increased within a few minutes. This was especially noticeable when TCEP was added. The observed light scattering is likely due to polymerization or formation of oligomers of Rv2633c. Cloudiness was visibly present in the protein. This was more significant with the addition of TCEP, where visual inspection of the solution showed what appeared to be precipitates.

Figure 30: Addition of 10 mM Glutathione causes Light Scattering that cannot be Reversed by Dialysis
After overnight dialysis in 50 mM tris pH 8 the protein still showed significant light scattering, but the cofactor appeared to no longer be present in the protein. Though hard to tell by UV-Visible absorbance spectrum due to the extreme light scattering, the protein color changed from light pink to almost completely clear. The reducing agents may have destabilized or unfolded the protein in some way, causing loss of cofactor. Due to the high concentrations
needed for noticeable change, the results may also have been caused by denaturation of the protein.

CONCLUSIONS

Sequence Analysis and Structural Modeling Predictions

Sequence analysis of Rv2633c shows very low homology with non-mycobacterium species. And within the mycobacteria, the presence of a protein sequence highly similar to Rv2633c correlates within the pathogenic strains of mycobacterium. Only one nonpathogenic strain of mycobacterium, *Indicus pranii*, is selected by NCBI Blast as a significant alignment, and it is predicted to be the furthest related mycobacterium species. The C terminal sequence also seems to be even more predictive of pathogenic species of mycobacteria.

Analysis of the predicted structural homology of Rv2633c proves difficult, due to the low sequence homology with available structural templates. However, there were several observations that could be drawn from the generated structural models. Each model produced had a similar overall structure of two sets of approximately parallel alpha helices arranged into a tunnel with a hollow center. The center of the helices is partially blocked by a fifth alpha helix on the carboxyl end. This general structure coincides with the known structure of hemerythrin and hemerythrin-like proteins. This evidence strongly suggests this as a good model for the general structure of Rv2633c.

Though the structural prediction of manganese binding Hypothetical protein NMB1532 was valued as the closest to the native structure of Rv2633c, this is not likely enough information to rule out the presence of the cofactor to be a diiron µ-oxo bridge. Though the NMB1532 model was rated the highest predicted model in overall structure, the predictions that include the
binding site of Rv2633c are actually rated lower in probability compared to models off of protein structures known to bind diiron. The length of sequence coverage is highly factored into probability calculations of predicted models, yet the extra amino acid residues that are included when using the NMB1532 model do not coincide with any predicted binding site.

**Purification and Expression**

Expressing large quantities of pure recombinant protein was fairly simple, and different expression and purification methods produced amounts between 2 and 3 mg/ml. The largest yield obtained was 59.8 mg from a 4 liter growth of cells. However, due to the unstable nature of the predicted cofactor binding, expression and purification of a stable, cofactor containing form of Rv2633c proved difficult. Purification using imidazole was unpredictable in terms of cofactor yield. All but one preparation resulted in no spectrally visible 330 nm peak. Purification with the EDTA strategy showed consistency in the presence of the cofactor, though at a lower apparent level. The 330 nm peak predicted to correspond to diiron cofactor was visually present, and the protein was reactive to hydrogen peroxide. However, when compared to the successful purification of cofactor containing protein using imidazole the peak is much less intense. Various expression and purification methods showed that the cofactor is easily destabilized by environmental conditions. Purification of histidine tagged Rv2633c with EDTA to strip the cobalt and protein off the column served as the most consistent method for purification, though the metal chelating properties likely destabilized a portion of the protein. In future studies, it may be more useful to clone Rv2633c into a new vector containing a different purification tag, such as a GST tag.
**In vitro Macrophage Conditions**

Based on experiments where oxygen levels were removed, or SNAP was used to generate nitric oxide, it is unlikely that Rv2633c reversibly binds dioxygen. However, whether Rv2633c is unable to bind O₂, or is bound too tightly to release under these conditions, cannot be determined. Neither of these methods commonly used to test for the binding of dioxygen showed any spectral changes. While it is possible that dioxygen binding occurs and is just not shown by changes in UV-visible absorbance, this is likely not the case as hemerythrin domains in varying protein and organisms show changes in absorbance corresponding to the presence and absence of dioxygen [42].

Rv2633c responded poorly to reducing conditions, and there was noticeable aggregation of protein that could not be reversed by dialysis. Prolonged exposure to reducing agents also seemed to strip the protein of the cofactor, as the 330 nm absorbance peak diminishes almost completely.

The highly oxidative environment generated by high concentrations of hydrogen peroxide caused considerable changes in Rv2633c. There appears to be two reactions occurring. The peak absorbance at 330 nm increases rapidly immediately upon addition of hydrogen peroxide, with small growth of a 550 nm absorbance peak. At approximately four minutes after hydrogen peroxide addition there is a rapid growth in the 550 nm peak, accompanied by the visual color change from light pink to purple. This may be due to changes in oxidation state of the two iron atoms predicted to be present in the protein center, as this reaction is almost completely reversible upon dialysis. The addition of the same concentration of tert-butyl hydrogen peroxide does not cause any noticeable shift in absorbance, therefore the hydrogen peroxide group likely must be able to enter the protein center in order to cause oxidation effects.
DISCUSSION

Tuberculosis infection continues to be a serious health risk around the world, especially for individuals in developing countries and those who are immunocompromised. With the continuing prevalence of Mtb antibiotic resistance, research into the virulence and persistence of Mtb is greatly needed. Rv2633c is a promising candidate for further investigation. The presence of this gene in mycobacteria strongly correlates with pathogenic species of mycobacteria. Rv2633c has been shown experimentally to be upregulated during the initial stages of virulence.

Rv2633c shows reversible reactivity with hydrogen peroxide. Based on the known literature of hemerythrin oxidation states, I predict the spectral shifts caused by hydrogen peroxide are caused by the µ-oxo diiron being oxidized from an Fe(II, II) or possibly (Fe II, III) state to a fully oxidized Fe (III, III) form. Oxidized hemerythrin has a very similar absorbance spectrum: no noticeable absorbance spectrum in the deoxy and diiron free form, a 330 nm peak observed in the met-form, and a 330 nm and a noticeable peak at around 550 nm only in the oxidized state [75].

My speculative prediction for the function of Rv2633c is iron sensing. Sequence analysis of Rv2633c predicts two different domains involved in the binding of iron: hemerythrin and ferritin-like domain. Experiments performed on the recombinant protein do not show any noticeable reversible dioxygen binding so the hemerythrin-like domain of Rv2633c is most likely not used to bind dioxygen. Based on experiments performed in this work, Rv2633c is likely not involved in iron storage. Mycobacteria have known proteins highly successful in iron sequestration and storage. The predicted diiron cofactor also experimentally seems to be bound unstably, so continual iron binding required for storage seems unlikely. Structural predictions are also unable to fully conserve cofactor binding sites. FBox and Leucine-Rich repeat protein
(FBXL5), one of the templates used for previous homology models of Rv2633c, has 20% sequence identity with Rv2633c, and has been characterized as an iron homeostasis protein [76]. FBXL5 senses cellular iron levels by binding μ-oxo diiron with a hemerythrin-like domain and destabilizing when not bound to iron [76]. Preliminary data from Dr. Kyle Rohde shows *Mtb* without the Rv2633c gene seem to grow better when compared to the wild type. Perhaps Rv2633c plays a role in sensing unfavorable conditions for growth.

Future studies of Rv2633c are needed to fully understand the part it plays in the virulence and survival of *Mtb* in the human macrophage. Determination of the true metal content of Rv2633c by ICP-MS would help elucidate the chemistry occurring when reacting in oxidizing environments. Further testing to determine if Rv2633c exists in its native state as a monomer or an oligomer would also give insight into its function. Various experimental conditions showed evidence of oligomerization by Rv2633c. SDS-PAGE of highly purified recombinant protein show recurrent faint bands at positions of high molecular weights. The addition of reducing agents caused fast and varying aggregation, which could possibly be due to the propensity to form oligomers. Hemerythrin proteins are known to form oligomers, and ferritin-like proteins are also shown to form extremely large oligomers, such as dodecamers, in order to function properly [42, 77]. A crystal structure is the next logical step for truly understanding the structure of Rv2633c.

Continued research into the function of Rv2633c in *Mtb* virulence will likely yield a greater understanding of the remarkable resilience of *Mtb* within the human macrophage. The high expression of this protein during uptake into the macrophage and acidification of the lysosome could prove a promising drug target in the fight against tuberculosis infection.
CITATIONS


