A Thesis

entitled

PhoR, PhoP and MshC: Three essential proteins of *Mycobacterium tuberculosis*

by

Erica Loney

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the

Master of Science Degree in Chemistry

________________________________________
Dr. Donald R. Ronning, Committee Chair

________________________________________
Dr. John J. Bellizzi, Committee Member

________________________________________
Dr. Ronald Viola, Committee Member

________________________________________
Dr. Patricia R. Komuniecki, Dean
College of Graduate Studies

The University of Toledo

May 2014
The tuberculosis (TB) pandemic is responsible for 1.6 million deaths annually, most of which occur in developing nations. TB is treatable, though patient non-compliance, co-infection with HIV, and the long, 6-9 month treatment regimen have resulted in the emergence of drug-resistant TB. For these reasons, the development of novel anti-tuberculin drugs is essential. Three proteins – PhoR, PhoP, and MshC – of *Mycobacterium tuberculosis* (*M.tb*), the causative agent of TB, are the focus of this thesis.

The PhoPR two-component system is a phosphorelay system responsible for the virulence of *M.tb*. The histidine kinase PhoR responds to a yet-unknown environmental stimulus and autophosphorylates a conserved histidine. The phosphate is transferred to an aspartate of the response regulator PhoP, which then forms a head-to-head homodimer and initiates the transcription of 114 virulence genes. Inhibiting PhoR or PhoP would rendered *M.tb* avirulent and increase its susceptibility to other drugs. For our PhoPR project, PhoR was separated into its extracellular (PhoRex) and intracellular (PhoRin) domains. We attempted to crystallize PhoRex and PhoRin, though currently no crystals
have been achieved. Using the malachite green assay PhoRin was confirmed to spontaneously dephosphorylate. The DNA binding activity of PhoP was assessed using fluorescence polarization, though no binding was observed. To observe the complexation of PhoRin and PhoP, a pull-down method was developed using the non-hydrolyzable ATP analog, adenosine 5′-(β,γ-imido)triphosphate (AMP-PNP). A stable PhoRin-AMP-PNP-PhoP complex was not identified. Future work is necessary for the crystallization of PhoRex and PhoRin. A new DNA binding assay is also necessary to elucidate PhoP kinetics. Obtaining these data will give significant insight to potential inhibitors for these proteins.

Mycothiol is the low molecular weight thiol responsible for protection against oxidative stress and electrophilic toxins in actinomycetes. Previous studies suggest MSH is critical for the survival of M.tb, leading to its attractiveness as a potential drug target. MshC is the penultimate enzyme of the mycothiol (MSH) biosynthetic pathway, catalyzing the ATP-dependent ligation of cysteine and glucosamine inositol in a Bi Uni Uni Bi Ping Pong mechanism. An incomplete 1.6 Å crystal structure of M. smegmatis MshC has previously been solved using trypsin digestion and a cysteine-adenylate analog, 5-O-[N-(L-cysteinyl)-sulfamonyl]adenosine (CSA). Steady state and pre-steady state kinetics for M. smegmatis MshC have also been determined. For this project, we attempted to crystallize the full-length M. smegmatis MshC using K_m values of ATP, L-cysteine, and the potential inhibitor glucosazide inositol. Thus far, crystallization attempts have not been successful. We also attempted to develop a high-throughput assay for drug screening. Using the quinaldine red assay, the specific activity of MshC was determined to be 2.40x10^{-2} µmol/min·mg. While the high-throughput assay is not yet
complete, the foundation has been developed for future optimization and usage. As a final goal, we desired to express \textit{M.\textit{tb}} MshC in \textit{E. coli} codon-optimized competent cells. This venture has not yet proved successful. Expression of \textit{M.\textit{tb}} MshC will allow for crystallization attempts of this protein as well as the determination of kinetic data.
Acknowledgements

I first wish to express my sincere gratitude to my advisor, Dr. Donald R. Ronning, for his wisdom, kindness, and limitless patience throughout my research journey. Thank you for being such a great mentor and for teaching me so much. The experience I have gained in your lab has been and will be significant for my future development as a scientist. I truly appreciate your guidance.

I would like to extend my gratitude to my committee members, Dr. John J. Bellizzi and Dr. Ronald Viola for their endless support and advice while I prepared my thesis. I could not have chosen a better committee!

To my labmates, both past and present, thank you so much for sharing your knowledge and for being such an awesome team to work with! I had a wonderful time with everyone. I would like to specifically acknowledge Vidhi, Lorenza, and Jared for showing me the ropes and keeping me saturated with valuable advice. I would also like to thank Dr. Peter Andreana and Krishnakant Patel for their contributions, both conceptually and synthetically, to the MshC project. Thank you to Jen, Steven, and Rachel for providing levity and endless support. Your friendship is invaluable.

Finally, to all my family and friends outside the university, thank you for seeing me through this journey and putting up with me during the rough lab days and long lab nights. To my mom and dad, my brothers Matthew and Charles, and my partner Clay, thank you from the bottom of my heart. I could not have done this without all of you.
Table of Contents

Abstract ........................................................................................................................................... iii

Acknowledgements ......................................................................................................................... vi

Table of Contents .......................................................................................................................... vii

List of Tables ..................................................................................................................................... x

List of Figures ..................................................................................................................................... xi

List of Equations ............................................................................................................................... xii

List of Abbreviations ....................................................................................................................... xiv

List of Symbols ............................................................................................................................... xvi

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

1.1 The Tuberculosis Burden ............................................................................................................. 1

1.1.1 Introduction to the Tuberculosis infectious process ............................................................... 1

1.1.2 History of treatment regimens .............................................................................................. 2

1.1.2.1 The BCG vaccine ............................................................................................................. 2

1.1.2.2 Development of first and second line drugs .................................................................. 3

1.1.2.3 Co-infection with HIV and emergence of drug resistance .............................................. 3

1.2 Roles of the PhoPR two-component system in Mycobacterium tuberculosis ......................... 4

1.1.1 Introduction to two-component systems ................................................................................. 4

1.1.2 PhoR: histidine kinase ......................................................................................................... 5

1.1.3 PhoR: current structural knowledge ..................................................................................... 5
1.1.4 PhoP: response regulator ...............................................................6
1.1.5 PhoP: current structural knowledge ............................................7

1.3 Roles of MshC in the mycothiol biosynthetic pathway of *Mycobacterium tuberculosis* .............................11

1.3.1 Mycothiol biosynthetic pathway ...............................................11
1.3.2 MshC: cysteine ligase ...............................................................13
1.3.3 MshC: current structural knowledge ........................................14

1.4 Project goals ..................................................................................17

1.4.1 PhoPR project goals .................................................................17
1.4.2 MshC project goals .................................................................17

2 Materials and Methods .......................................................................19

2.1 Molecular cloning, expression, and purification of PhoP, PhoR, and MshC ..19

2.1.1 Molecular cloning ....................................................................19
2.1.2 Protein expression ....................................................................22
2.1.3 Protein purification ....................................................................24

2.2 Mass spectrometry analyses ...........................................................27

2.2.1 MALDI: peptide fingerprinting .................................................27
2.2.2 ESI .........................................................................................28

2.3 Activity assays .............................................................................29

2.3.1 Malachite Green assay ............................................................29
2.3.2 Quinaldine Red assay ...............................................................30
2.3.3 Fluorescent Polarization assay ................................................33
2.3.4 Pull-Down assay .....................................................................33
2.4 Crystallization strategy.................................................................34

3 PhoPR Results..................................................................................35
  3.1 Expression and purification of PhoP and PhoR...............................35
  3.2 MALDI peptide fingerprinting.....................................................37
  3.3 Malachite green activity assays..................................................38
  3.4 Fluorescent polarization activity assay.........................................39
  3.5 PhoPR complex...........................................................................39
  3.6 Crystallization attempts.............................................................40
  3.7 Discussion..................................................................................43

4 MshC Results..................................................................................47
  4.1 Expression and purification of MshC...........................................47
  4.2 Confirmation of product formation using ESI-MS..........................49
  4.3 Mycobacterium smegmatis MshC: quinaldine red assay.................50
  4.4 Crystallization attempts.............................................................52
  4.5 Discussion..................................................................................52
    4.5.1 Discussion of results............................................................52
    4.5.2 Current enzyme kinetics.......................................................54

5 Conclusions and future work..........................................................58

References..........................................................................................61

Appendix A Sequences of phoR and PhoR...........................................67
Appendix B Sequences of phoP and PhoP.............................................69
Appendix C Sequences of M. smegmatis and M.tb mshC and MshC........70
List of Tables

3.1 Sequences of forward and reverse primers for PhoP, PhoRex, and PhoRin........36
3.2 Malachite green assay results.................................................................39
3.3 Results of PhoP-DNA fluorescence polarization assay..........................40
4.1 Sequences of forward and reverse primers for M. smegmatis and M.tb MshC.....48
# List of Figures

1-1 Cartoon of EnvZ, a model for the cytoplasmic domain of histidine kinases.........6
1-2 Cartoon of the PhoP dimer.................................................................................8
1-3 Phosphorylation site of PhoP.............................................................................9
1-4 Cartoon of the PhoB DNA-binding domain overlaid with PhoP....................10
1-5 Mycothiol biosynthetic pathway....................................................................12
1-6 Sequence alignment of \textit{M.tb} and \textit{M. smegmatis} MshC.............................14
1-7 Cartoon of \textit{M. smegmatis} MshC................................................................16
1-8 Illustration featuring the interactions of CSA with MshC active site residues....17
2-1 Step-by-step Gibson assembly scheme..............................................................23
2-2 Malachite green assay scheme.........................................................................30
2-3 A potential inhibitor of MshC, glucosazide inositol.........................................32
2-4 Quinaldine red assay scheme............................................................................33
3-1 Example of a PhoP nickel affinity chromatogram featuring a step gradient......37
3-2 PhoPR bands from SDS-PAGE after different stages of purification................38
3-3 Phosphate standard curve for MG assay............................................................39
3-4 SDS-PAGE gel of the pull down method............................................................41
3-5 Samples of PhoRex crystals................................................................................43
4-1 Example of an MshC cobalt affinity chromatogram..........................................49
4-2 \textit{M. smegmatis} MshC bands on SDS-PAGE gel after various purification steps..49
<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-3</td>
<td>ESI spectra indicating the protonated mass of Cys-GlcN-Ins</td>
<td>50</td>
</tr>
<tr>
<td>4-4</td>
<td>QR endpoint MshC assay results</td>
<td>51</td>
</tr>
<tr>
<td>4-5</td>
<td>QR assay varying the concentration of MshC</td>
<td>52</td>
</tr>
<tr>
<td>4-6</td>
<td>Phosphate standard curve for the QR assay</td>
<td>52</td>
</tr>
<tr>
<td>4-7</td>
<td>Schematic of the MshC Bi Uni Uni Bi Ping Pong mechanism</td>
<td>56</td>
</tr>
</tbody>
</table>
List of Equations

2-1 Beer-Lambert law ........................................................................................................27
List of Abbreviations

AcCys-GlcN-Ins.................Acetylcysteine glucosamine inositol
AMP-PNP........................Adenosine 5′-(β,γ-imido)triphosphate
ATP-γ-S..........................Adenosine 5′-[γ-thio]triphosphate

BCG.................................Bacille-Calmette-Guérin
BME.................................β-mercaptoethanol
B. subtilis.........................Bacillus subtilis

CDC...............................Center for Disease Control and Prevention
CHCA.............................α-Cyano-4-hydroxycinnamic acid
Cys-GlcN-Ins.....................Cysteine glucosamine inositol
CysRS..............................Cysteinyl-tRNA synthetase

E. coli....................Escherichia coli
ESI...............................Electrospray ionization

FP.................................Fluorescence polarization
FPLC..............................Fast Protein Liquid Chromatography

GFP...............................Green fluorescent protein
GlcAz-Ins..........................Glucosazide inositol
GlcN-Ins..........................Glucosamine inositol
GST.................................Glutathione S-transferase

HIV...............................Human Immunodeficiency Virus
HK.................................Histidine kinase

IPTG..............................Isopropyl β-D-1-thiogalactopyranoside

MALDI............................Matrix-assisted laser desorption/ionization
M. bovis.........................Mycobacterium bovis
MBP.................................Maltose binding protein
MG.................................Malachite green
MSH.................................Mycothiol
M. smegmatis.....................Mycobacterium smegmatis
M. tb...............................Mycobacterium tuberculosis
PhoRex..........................Extracellular domain of PhoR
PhoRin..........................Intracellular domain of PhoR

QR..............................Quinaldine red

RR..............................Response regulator

TB..............................Tuberculosis (disease)
TCS..............................Two-component system
TFA..............................Trifluoroacetic acid
TOF..............................Time-of-flight
TRX..............................Thioredoxin

WHO..............................World Health Organization
List of Symbols

Å.........................Angstrom
Ax..........................Absorbance at x nanometers
Chapter 1

Introduction

1.1 The Tuberculosis Burden

1.1.1 Introduction to the Tuberculosis infectious process

Tuberculosis (TB) is an infectious disease caused by the *Mycobacterium tuberculosis* (*M.tb*) bacillus. According to the World Health Organization, nearly one-third of the world’s population is infected with TB\(^1\). In 2012, 8.6 million new cases of TB developed, and 1.3 million people died from the disease\(^1\). TB is one of the leading causes of death by an infectious agent, coming second only to the Human Immunodeficiency Virus (HIV)\(^1\).

*M.tb* is an airborne pathogen that is transmitted through aerosolization\(^2\). It infects alveolar macrophages and dendritic cells within the lungs\(^2\). This mycobacterium has evolved impressive survival mechanisms that contribute to its endurance in the digestive conditions of the macrophage phagosome. Several mechanisms of note include reduction of phagosome acidity, prevention of phagosome-lysosome fusion, and the presence of a uniquely complex bacterial cell wall\(^2\). Infected persons may experience either a latent or active infection. The latent infection is asymptomatic, while most active infections feature a productive cough with sputum and blood, fever, night sweats, weakness, chest
pain, weight loss, and formation of granulomas in the lungs (pulmonary TB)\(^2\). In 10-20% of active infections, the granulomas form in areas other than the lungs, and this variation is considered extrapulmonary TB\(^2\). Many of the deaths caused by this disease are preventable\(^1\).

**1.1.2 History of Treatment Regimens**

**1.1.2.1 The BCG vaccine**

In 1908, Albert Calmette and Camille Guérin began developing a vaccine against TB using a live attenuated strain of *Mycobacterium bovis* (*M. bovis*)\(^3\). This bacillus is related to *M.tb* and is the causative agent of bovine tuberculosis, although it can also cause human TB if ingested\(^4\).

In the development of this vaccine, it was observed that a strain of *M. bovis* cultivated in glycerin-bile-potato media exhibited lower virulence than typical\(^3\). Over a period of years, continuous serial cultivation of *M. bovis* in this media resulted in spontaneous attenuation of the live bacillus\(^3\). This strain was then used to create the Bacillus-Calmette-Guérin (BCG) vaccine, which was first approved for human usage in 1921\(^3\). Though uncommon in the Western World, many developing nations still utilize the BCG vaccine to this day\(^3\).

The BCG vaccine is the only anti-tuberculin vaccination currently approved for human use, but its effectiveness is variable. Depending on the region, protection ranges from 0-80\%\(^3\). Decades of cultivation have produced substrains of *M. bovis* genetically modified to the extent that they have lost their immunogenicity\(^3\). Other factors such as diversity in human genetics and previous exposure to environmental mycobacteria also
contribute to the vaccine’s lack of efficacy. Because a high-fidelity vaccination is not currently available, other methods of protection against TB are necessary.

### 1.1.2.2 Development of first and second line drugs

There are four first line drugs involved in treating TB. Ethambutol (1961) inhibits lipid biosynthesis and isoniazid (1952) inhibits the fatty acid synthesis. Pyrazinamide (1952) was thought to also inhibit fatty acid biosynthesis, although new developments suggest this is not the correct mechanism of action. Finally, rifampicin (1957) inhibits RNA polymerase. TB is generally managed with a combination of these four drugs, and treatment typically lasts 6-9 months.

Second line drugs become necessary when the disease does not respond to first line treatment. Second line drugs tend to be less effective, less accessible, and/or cause more severe side effects. Some examples of these drugs include kanamycin, capreomycin, ethionamide, closerin, and ciprofloxacin.

Over fifty years have passed since the four first line drugs have been developed. Significant progress in new anti-tubercular medications must be made to combat the TB epidemic still plaguing the world today.

### 1.1.2.3 Co-infection with HIV and emergence of drug resistance

Tuberculosis is a treatable disease. However, its prevalence indicates that eradicating the disease is very complicated. TB is most common in developing nations, with high burden areas such as India, Southeast Asia, and Africa accounting for 80% of cases. Developing nations also have a higher incidence of HIV infection, and HIV-positive persons are more susceptible to TB infection. This presents a conundrum, as the
drugs needed to treat HIV and TB are contraindicated\(^1\). Thus, the development of new drugs is essential for tackling these cases.

As mentioned previously, TB treatment spans 6-9 months. Because many high burden countries are also developing nations, access to healthcare, health education, and affordable drug products can be limited. Patient noncompliance, whether voluntary or through an inability to continue treatment, has lead to the emergence of drug resistant TB. Multidrug Resistant TB (MDR-TB) is resistant to rifampicin and isoniazid, the two most potent first line drugs\(^1\). Extensively Drug-Resistant TB (XDR-TB) is resistant to all first line drugs and some second line drugs\(^1\). Recently, cases of Totally Drug-Resistant TB (TDR-TB) have been reported in India; TDR-TB is resistant to all known anti-tuberculin drugs\(^1\). The emergence of such widespread drug resistance further emphasizes the need for new drug development.

1.2 Roles of the PhoPR Two-Component System in *Mycobacterium tuberculosis*

1.1.1 Introduction to two-component systems

Two-component systems (TCS) are regulatory mechanisms employed by bacteria for sensing and responding to environmental conditions\(^{10}\). A TCS is composed of a membrane-bound histidine kinase (HK) coupled with a response regulator protein (RR)\(^{10}\). There are many different TCSs, but each follows a similar mechanism. First, the HK is exposed to an environmental stimulus – this could be in the form of a ligand, an ion, or a change in pH, for example\(^{10}\). When this periplasmic domain is stimulated, a conserved histidine of the HK’s cytoplasmic domain is phosphorylated using ATP\(^{10}\). After this autophosphorylation reaction, the phosphoryl group is then transferred to a conserved
aspartate of the RR\textsuperscript{10}. This stimulates the RR to express genes appropriate for the current environmental conditions\textsuperscript{10}. Our TCS of interest is the PhoR-PhoP system of \textit{Mycobacterium tuberculosis}.

1.1.2 PhoR: histidine kinase

PhoR is the transmembrane HK of this TCS. The ligand/condition that PhoR senses is currently unknown. In similar organisms, PhoR homologs sense phosphate\textsuperscript{11} (\textit{Bacillus subtilis}) or Mg\textsuperscript{2+} and antibacterial peptides\textsuperscript{12} (\textit{Salmonella enterica typhimurium}). The mechanism through which PhoR is autophosphorylated is unknown. Though a majority of HKs are known to form homodimers that undergo trans-phosphorylation\textsuperscript{13}, this phenomenon is not observed in \textit{M.tb} PhoR.

1.1.3 PhoR: current structural knowledge

PhoR is approximately 52,000 Daltons in size. No structural information is available. While crystal structures of homologous HKs have been solved, the periplasmic sensing domains vary greatly in their structures due to the individuality of their sensory needs\textsuperscript{13}. However, the cytoplasmic domains appear to share conserved structural motifs. These motifs are referred to as the H, N, G1, F, and G2 boxes\textsuperscript{13}. These highly flexible, hinge-like loops are ordered around a conserved \(\alpha-\beta\) sandwich to form the ATP-binding pocket\textsuperscript{13}. The \(\alpha-\beta\) sandwich is composed of mixed \(\beta\)-sheets surrounded by \(\alpha\)-helices, with the number of sheets and helices varying between HK species\textsuperscript{13}. The N, G1, and G2 boxes plus an essential Mg\textsuperscript{2+} ion coordinate the nucleotide, and G2 is predicted to act as a lid that closes over the inhabited active site\textsuperscript{13}.
1.1.4 PhoP: response regulator

PhoP is the RR of the *M.tb* TCS. When an active site aspartate is phosphorylated by the HK, PhoP forms a homodimer. This alters the expression of 114 virulence genes: 44 genes are up-regulated and 70 are down-regulated. Included in this expression cascade are genes responsible for lipid metabolism, essential for formation of the cellular envelope. The notion that PhoP is directly responsible for the pathogenicity of this mycobacterium is supported by the existence of the avirulent H37Ra strain of *M.tb*. In this attenuated variant, PhoP exhibits a point mutation of a serine into a leucine in its
DNA binding domain.\textsuperscript{15} When the \textit{phoP} gene is restored in H37Ra, its virulence returns.\textsuperscript{15} Thus, PhoP – and by extension, PhoR – presents an attractive target for drug development, as inhibition of any point along this TCS pathway will result in an avirulent mycobacterium with a weakened cellular envelope.

\textbf{1.1.5 PhoP: current structural knowledge}

PhoP is approximately 28,000 Daltons in size. The crystal structure of this protein was solved by Menon and coworkers. PhoP belongs to the OmpR/PhoB subfamily, members of which are characterized by an N-terminal receiver domain and a C-terminal DNA-binding (effector) domain\textsuperscript{14}. The receiver domain is composed of 5-stranded $\beta$-sheet, surrounded on both sides by $\alpha$-helices\textsuperscript{14}. The DNA-binding domain is composed of three $\alpha$-helices and two $\beta$-sheets in a formation known as winged helix-turn-helix\textsuperscript{14}. Both domains exhibit structures that are conserved within the PhoB subfamily\textsuperscript{14}. 

PhoP forms a head-to-head homodimer through its receiver domain\textsuperscript{14}. The two monomers interact symmetrically through their $\alpha 4$-$\beta 5$-$\alpha 5$ faces, as seen above in Figure 1.2\textsuperscript{14}. The two DNA-binding domains do not interact. The cognate PhoR transfers a phosphate group to Asp71 of the PhoP receiver domain\textsuperscript{14}. How this results in the activity...
of the DNA-binding domain is still unknown, though it is possible that phosphorylation stabilizes the dimeric form.

It is also unknown how PhoP binds DNA. However, it is likely that PhoP binds DNA in a fashion similar to the DNA-binding domain of PhoB, a homolog found in *E. coli*. The structure of PhoB bound with DNA has been solved, and by superimposing the PhoB-DNA complex with PhoP, many similarities are evident. The terminal residues of helix α7 can interact with the sugar and phosphate backbone, as can N-terminal residues Pro176 and Thr177, which are conserved between PhoP and PhoB. Ser219 of PhoP would appear to hydrogen bond with the DNA backbone in the overlapping model. This interaction is apparently significant, as the PhoP of the H37Ra strain exhibits a point mutation of Ser219 to a leucine. The leucine would not only be unable to hydrogen bond, but it would also cause steric hindrance due to its bulkier side chain. This would

Figure 1.3 Phosphorylation site of PhoP, including relevant residues. This site is located in a shallow pocket near the C-terminal. Hydrogen bonds are indicated with dashed lines. Asp71 is the site of phosphorylation. PDB code: 3R0J
inhibit the PhoP-DNA complex, affecting gene expression and ultimately resulting in the avirulence of H37Ra strain.

However, the PhoB-DNA complex is not a perfect model for the PhoP DNA-binding domain. Most notably, PhoB operates as a tandem dimer. PhoP is unlikely to experience tandem dimerization, as the interfaces necessary for such association would prevent DNA binding. Tandem dimerization would recruit helices $\alpha 6$ and $\alpha 8$; as these same regions participate in DNA binding according to Wang and coworkers’ model, the recognition helix of PhoP would be unable to insert into the DNA major groove.

Figure 1.4  Cartoon of the PhoB DNA-binding domain (pink) overlaid with PhoP (green). PDB code: 1GXP
1.3 Roles of MshC in the mycothiol biosynthetic pathway of *Mycobacterium tuberculosis*

1.3.1 Mycothiol biosynthetic pathway

*Mycobacterium tuberculosis* produces mycothiol (acetylcysteine glucosamine inositol, AcCys-GlcN-Ins or MSH) to protect against oxidative stress, electrophilic toxins, and antibiotics\(^\text{17}\). This pathway is unique to actinobacteria\(^\text{17}\), which attracts attention in the search for new drug targets.

There are five enzymes involved in the mycothiol biosynthetic pathway. First, *N*-acetyl glucosamine transferase (MshA) ligates an acetylated glucosamine to a phosphoinositol to form 3-phospho-GlcNAc-Ins\(^\text{17}\). An unidentified phosphatase dephosphorylates this molecule to yield GlcNAc-Ins, which is then deacetylated by MshB to yield GlcN-Ins\(^\text{17}\). Cysteine ligase (MshC) ligates GlcN-Ins with cysteine in an ATP-dependent reaction\(^\text{17}\). In the final step, MshD acetylates Cys-GlcN-Ins to produce MSH\(^\text{17}\).
Strains of *M. smegmatis* were selected for mutations involving the genes corresponding to *mshA*-D in *M.tbc* to determine the significance of each resultant enzyme. MSMEG_0933 (*mshA* homologue) mutant presented an increased resistance to isoniazid, along with increased sensitivity to hydrogen peroxide and rifampicin. The *mshB* mutant experienced a 90% decrease in the production of MSH, though no significant increase of antibiotic or oxidative sensitivity was observed. The *cysS* (*mshC* homologue) mutant became extremely susceptible to oxidative stress, though it was more resistant to many antibiotics such as ethionamide and isoniazid. Finally, the *mshD* mutant was similar to the wild type in terms of susceptibility and resistance, as this mutant produced novel...
thiols (\(N\)-formyl-Cys-GlcN-Ins and \(N\)-succinyl-Cys-GlcN-Ins) which seem to function similarly to MSH\(^{17}\).

\(MshA-D\) appear to be non-essential in \(M.\) \textit{smegmatis}. However, in \(M.tb\), \(mshC\) was observed to be essential for in vitro growth\(^{18}\). Considering \(M.tb\) grows six times slower than \(M.\) \textit{smegmatis}, it is plausible that \(M.\) \textit{smegmatis} can outgrow the rate of lethal oxidative damage without MSH, while \(M.tb\) cannot\(^{18}\). Thus, MshC presents an attractive target for drug development.

1.3.2 MshC: cysteine ligase

MshC catalyzes the penultimate step in mycothiol biosynthesis. Cysteine is ligated to GlcN-Ins in a mechanism following a Bi Uni Uni Bi Ping Pong schematic\(^{18}\). In this mechanism, ATP binds to MshC first, followed by the binding of cysteine to form a ternary Enzyme-ATP-cysteine complex\(^{18}\). MshC releases pyrophosphate when this complex is formed\(^{18}\). Next, GlcN-Ins binds to MshC and the adenylate anhydride is cleaved, forming an amide bond between GlcN-Ins and cysteine\(^{18}\). Finally, Cys-GlcN-Ins is released from the active site along with AMP, though the order of product release has not been identified\(^{18}\).

MshC shares a common evolutionary origin with cysteinyl-tRNA synthetase (CysRS)\(^{19}\). Sequence comparisons between CysRS and MshC indicate a 36% similarity\(^{19}\). MshC is highly conserved between \(M.tb\) and \(M.\) \textit{smegmatis}, with a sequence similarity over 80%.
1.3.3 MshC: current structural knowledge

MshC is approximately 47,000 Daltons in size. While the crystal structure of \( M. tb \) MshC has not yet been solved, Tremblay and coworkers solved a 1.6 Å crystal structure of MshC from \( M. smegmatis \) in 2008. Considering the sequence similarity of these two homologs, it can be inferred that both versions of MshC have comparable three-dimensional structures and active sites.

The \( M. smegmatis \) MshC crystal structure was achieved by utilizing limited proteolysis to cleave flexible regions of the enzyme while leaving the core intact, meanwhile a cysteinyl adenylate analog, 5-O-[N-(L-cysteinyl)-sulfamonyl]adenosine (CSA) mimicked the reaction intermediate\(^{19}\). The three-dimensional structure of MshC is

Figure 1.6 Sequence alignment of \( M. tb \) and \( M. smegmatis \) MshC. Significant conserved regions include the KMSKS loop responsible for GlcN-Ins positioning (red square) and the active site residues indicated in blue highlights.
similar to that of CysRS. There are four domains: the anticodon binding domain, the connective polypeptide (CP) domain, the stem-contact (SC) fold and the Rossman fold\textsuperscript{19}.

The anticodon binding domain consists of a C-terminal antiparallel $\alpha$-helix domain\textsuperscript{19}. Its name is carried over from homologous domain of the CysRS structure – it should be noted that MshC does not exhibit anticodon binding activity. The CP domain consists of a four-stranded antiparallel $\beta$-sheet motif\textsuperscript{19}. This domain connects the two halves of the Rossman fold.\textsuperscript{19} The SC domain connects the C-terminal with the Rossman fold and contains a KMSKS motif located on a flexible loop\textsuperscript{19}. This loop is not visible in the crystal structure, as it was cleaved during proteolysis\textsuperscript{19}. The Rossman fold contains the active site, and is composed of a five-stranded $\beta$-sheet surround by $\alpha$-helices\textsuperscript{19}. This domain is divided into two sections that are joined by the CP domain as mentioned previously.
In the active site, Zn$^{2+}$ coordinates the cysteine thiol$^{19}$. This zinc ion also coordinates the side chains of C43, C231, and H256 in a tetrahedral complex$^{19}$. Within the core of the Rossman fold, residues T46, T58, T83, G249, D251, and I283 bind the CSA ligand through a series of small loops$^{19}$. CSA is stabilized through four positions. First, the adenine ring inserts into a narrow pocket and experiences hydrogen bonding with T58 and I283$^{19}$. Secondly, the 2'-ribose hydroxyl hydrogen bonds with D251 and
G249\textsuperscript{19}. Though not observed in the CSA model, the highly conserved H44 is predicted to hydrogen bond with the endocyclic ribose oxygen of MshC’s true ligand (at 3.5 Å, H55 is not close enough to interact with CSA)\textsuperscript{19}. Third, the $\alpha$-amino group of the cysteine moiety hydrogen bonds with T46, T83, and G44\textsuperscript{19}. Finally, the fourth stabilization point consists of the previously discussed zinc-thiolate interaction\textsuperscript{19}.

The MshC active site has a structure that allows it to accept random binding of ATP and cysteine in the first half reaction; therefore, the order in which these substrates interact with MshC is not significant\textsuperscript{19}. The proteolyzed loop of the SC fold is predicted to both aid in the adenylate formation in the first half reaction, as well as orient the GlcN-Ins for the second half reaction\textsuperscript{19}. This is based on the function of the respective loop in

Figure 1.8 Illustration featuring the interactions of CSA with MshC active site residues.\textsuperscript{19}
CysRS, which positions the tRNA substrate. Based on kinetic assays featuring related aminoacyl-tRNA synthetases, the release of Cys-GlcN-Ins from the MshC active site is proposed to precede that of AMP. Considering the active site structure supports this hypothesis: GlcN-Ins is purportedly oriented by the SC loop after the formation of the cysadenylate. The adenine ring is nestled deep in the active site, with the newly formed Cys-GlcN-Ins lying in a shallower region. Therefore, it is logical that when the SC loop moves to allow the release of products, Cys-GlcN-Ins will leave the active site first as this product is located closer to the surface and the SC loop can no longer lock in the position of the GlcN-Ins moiety.

1.4 Project goals

1.4.1 PhoPR project goals

The PhoPR system is essential for \( M.tb \) virulence. Our goals for this project are to determine the crystal structures of the PhoR intracellular and extracellular domain, leading to the identification of the PhoR ligand. It is also desired to shed new light on the PhoPR phosphotransfer complex. After these aspects of PhoR and PhoP are elucidated, the identification of potential new drug molecules can commence.

1.4.2 MshC projects goals

We aim to complete the crystal structure of \( M. smegmatis \) MshC in order to observe the disordered loop regions that were previously proteolyzed. We also desire to develop a high throughput assay to allow for identifying potential drug candidates for MshC. In addition, another objective is determination of the crystal structure and kinetic
parameters of *M.tb* MshC. The development of novel drugs to target this vital enzyme in mycothiol biosynthesis is the ultimate goal.
Chapter 2

Materials and Methods

2.1 Molecular cloning, expression, and purification of PhoP, PhoR, and MshC

Our laboratories are not equipped for the handling of dangerous organisms such as M.tb. For this reason, we must employ molecular cloning techniques to insert our genes of interest into nonpathogenic E. coli so that we may express our protein of interest. Once expressed, our protein of interest must be separated from the rest of the cellular contents and we may then proceed with experimentation.

2.1.1 Molecular cloning

Our genes of interest were either synthesized by GeneArt® or cloned from the H37Rv strain of M.tb. The genes were amplified by touchdown PCR using primers ordered from Integrated DNA Technologies. PCR was performed using a thermocycler (Eppendorf), and for the touchdown sequence, the annealing temperature started at 65 °C and decreased 0.5 °C per cycle until a temperature of 60 °C was reached. The phor gene was cloned as two separate domains: the extracellular domain (PhoRex) and the intracellular domain (PhoRin). See Appendix A for the sequences of the separated domains. The remaining genes were cloned as their native forms. The forward and reverse primers each contained a restriction site. After the PCR is completed, the samples
were loaded onto a 1% agarose gel for gel electrophoresis to affirm the presence and purity of each gene. The bands corresponding to each gene were excised using a razor blade and weighed. The DNA was extracted from the gel using the materials and protocol provided by a QIAquick® Gel Extraction Kit, category number 28704. The eluate was stored at -20 °C.

After gel extraction, the genes encoding PhoP and the extracellular domain of PhoR were ligated between NdeI and HindIII in pDR28p. The gene encoding the intracellular domain of PhoR was ligated between NdeI and BamHI-HF. The gene encoding *M. smegmatis* MshC was ligated between NdeI and XhoI.

One Shot® Top10 chemically competent *E. coli* have been treated with calcium chloride to increase the porosity of their cell walls. Twenty-five microliters of Top10 cells were transformed with 1 µL of recombinant, non-supercoiled DNA and then incubated on ice for 30 minutes. The cells were then heat shocked in a 42 °C water bath for 30 seconds, causing the pores of the cell walls to relax and allow for the uptake of foreign DNA. The sample is then set back onto ice for two more minutes, allowing the pores to close and prevent DNA from leaking out. The Top10 cells are encouraged to enter the growth phase by the addition of 250 µL LB media and incubation at 37 °C for one hour. The sample is then plated, using aseptic technique, onto an agar dish infused with the antibiotic corresponding to the plasmid’s resistance and incubated overnight at 37 °C. If pET28 was used, 0.05 mg/mL kanamycin and 0.03 mg/mL chloramphenicol were the chosen antibiotics. If pET32 was used, 0.1 mg/mL carbenicillin was chosen.

If colonies grow overnight, it can be assumed that the *E. coli* have taken up the recombinant DNA. A portion of colonies are then selected and transferred to 14 mL
round bottom BD Falcon tubes, each containing 5 mL LB media plus the appropriate antibiotic. The cultures are incubated at 37 °C overnight, after which the cells are harvested by centrifuging at 12,000 rpm for 15 minutes. The plasmids are then extracted from the cells and purified using a QIAprep® Spin Miniprep Kit. The eluate was stored at -20 °C.

To confirm uptake of the target gene, the recombinant DNA is subjected to a restriction digestion. For this procedure, 10 units of each appropriate restriction endonuclease (determined by the identity of the vector and how the primers were engineered) were combined with 2.0 µL of the appropriate NEB buffer (determined by the endonucleases used) and mixed with 15 µL recombinant DNA. The sample is then incubated at 37 °C for one hour. After the digest is completed, the sample is loaded onto a 1% agarose gel. If the gene had been successfully ligated into the plasmid, one band corresponding to the plasmid and a second band corresponding to the size of the gene insert will be observed on the agarose gel. The purified plasmid DNA is stored at -20 °C. A 10 µL sample of each new construct is sent to Eurofins MWG Operon for sequence confirmation.

In some cases, molecular cloning was performed using Gibson assembly. Gibson assembly is a convenient method that ligates multiple DNA fragments together in a single step reaction. The Gibson assembly kit was purchased from New England BioLabs and the protocol was scaled down to a 5 µL reaction containing 0.5 µL PCR product, 0.5 µL linearized plasmid, 2.5 µL of the Gibson master mix, and 1.5 µL water. An overlap of 20 or more base pairs between the DNA fragments is required for this technique. The master mix contains exonuclease to cleave nucleotides from the 5’ end of the DNA (allowing the
resultant sticky ends to anneal), DNA polymerase to fill in the gaps, and DNA ligase to form phosphodiester bonds in the nicks of the backbone. The reaction mixture is incubated at 50 °C for 15 minutes. The product can then be used immediately for transformations.

Figure 2.1 Step-by-step Gibson assembly scheme. This reaction occurs at 50 °C and takes only 15 minutes, providing a convenient cloning option.

2.1.2 Protein expression

Once the sequence of the gene of interest was confirmed by nucleotide sequencing, *E. coli* competent cells for expression are transformed. These chemically competent cells are engineered to contain a T7 RNA polymerase gene while lacking the genes significant for lac operon activity. In the plasmid, a T7 promoter lies directly upstream of the inserted gene. Isopropyl β-D-1-thiogalactopyranoside (IPTG), an analog of the lactose metabolite allolactose, is used to induce the lac operon. This activates the
T7 promoter, which recruits T7 RNA polymerase from the transformed *E. coli*. The T7 RNA polymerase begins the transcription of genes located directly downstream. Unlike allolactose – which binds to the lac repressor and is metabolized by β-galactosidase – IPTG is non-hydrolyzable, so it remains bound to the lac repressor and promotes the continuous transcription of downstream genes.

Rosetta® or BL21* competent cells were transformed with recombinant, supercoiled DNA. The sample was then used to inoculate 5 mL LB media, to which the appropriate antibiotic was added. The culture incubated at 37 °C with agitation for three hours, or until turbid.

At this point, 2 mL of the uninduced cells were isolated and the rest of the culture was induced with 1 M IPTG for a test expression. The culture incubated at 37 °C for three hours with agitation. To prepare the sample for purification, the cells were harvested via centrifugation at 4000 rpm for 30 minutes, and resuspended in the appropriate binding buffer: 20 mM Tris pH 8.5, 100 mM NaCl, 25 mM imidazole, and 5 mM BME for nickel affinity, or 20 mM Tris pH 8.5 and 100 mM NaCl for cobalt affinity. The cells were treated with 1 mM DNAse and lysozyme, and incubated on ice for 30 minutes. The sample was then sonicated using a Misonix Sonicator 3000. Cellular debris was pelleted by centrifugation at 12,000 rpm for 15 to 30 minutes. The supernatant was purified using 100 μL Nickel NTA resin in a 1.5 mL microcentrifuge tube. To prepare the resin for batch protein purification, it was first washed with 1 mL deionized water and centrifuged for 5 minutes at 1000 rcf. The supernatant was pipetted off, and the resin was equilibrated with two washes of 1 mL equilibration buffer (20 mM Tris pH 8.5, 100 mM NaCl, 25 mM imidazole, and 5 mM BME) each. Each wash was centrifuged for 5
minutes at 1000 rcf and the supernatant was pipetted off. Once the resin was equilibrated, the analyte was added and incubated on ice for 15 minutes. After incubation, the sample was centrifuged for 5 minutes at 1000 rcf and the supernatant was reserved. The resin was washed with 1 mL equilibration buffer to remove any unbound protein. The sample was centrifuged for 5 minutes at 1000 rcf and the supernatant was reserved. Finally, 1 mL of elution buffer (20 mM Tris pH 8.5, 100 mM NaCl, 250 mM imidazole, and 5 mM BME) was added to elute bound protein. A final centrifugation step was performed and the supernatant reserved. Samples from the uninduced and induced unprocessed cells, along with samples from the wash and elution fractions of the metal affinity purification were loaded onto 12% to 15% polyacrylamide gels for SDS-PAGE. When the expression of soluble protein is achieved, the method can be scaled up to allow for the production of workable protein loads.

2.1.3 Protein purification

After confirming the expression of the target protein, the protein must be purified before any analyses can be performed. Metal affinity chromatography is a reliable method for protein purification. All of our plasmids contain genes encoding a poly-histidine tag at the N-terminal of the desired protein. The poly-histidine tag has a strong affinity for nickel and cobalt resins, and will bind exclusively to the metal stationary phase while the majority of other proteins and impurities simply wash through. Washing the column with a buffer of increased imidazole concentration can then elute the target protein. Imidazole has a stronger affinity for the metal ions, and out-competes the histidine tagged protein. All chromatography for this project was carried out utilizing Fast Protein Liquid Chromatography (FPLC) on an ÄKTA FPLC from Amersham.
Biosciences. Monitoring the absorbance at 280 nm allows for the easy identification of protein-containing fractions.

To start, 1 or 4 L of LB media were inoculated with transformed expression cells and the appropriate antibiotic(s). The culture was grown, with agitation, at 37 °C until reaching an OD$_{600}$ of approximately 0.7 to 0.8. One millimolar IPTG was added, and the culture was incubated with agitation at 16 °C for at least 16 hours. The cells were prepared for purification as described in Section 2.1.2. The supernatant, which contains all soluble protein, was filtered using a 0.45 µm syringe filter. A 1 or 5 mL HisTrap™ HP Ni or HiTrap™ TALON® crude Co affinity column from GE Healthcare was equilibrated with 5 column volumes of the appropriate binding buffer. The binding buffer contained 20 mM Tris pH 8.5, 100 mM NaCl, 25 mM imidazole, and 5 mM BME for nickel affinity, and 20 mM Tris pH 8.5 and 100 mM NaCl for cobalt affinity. The filtered supernatant was then loaded onto the column. The column was washed with five to ten column volumes of binding buffer to remove any unbound protein, and then an elution buffer was run through the column in either a stepwise or linear gradient fashion until all bound protein had been eluted. The elution buffer contained 20 mM Tris pH 8.5, 100 mM NaCl, 250 mM imidazole, and 5 mM BME for nickel affinity, and 20 mM Tris pH 8.5, 100 mM NaCl, and 150 mM imidazole for cobalt affinity.

If further purification was necessary, there are several options. Commonly, the poly-histidine tag would be removed. This can be significant for crystallization attempts, as the non-native tag may hinder crystal formation. To remove the tag, the fractions confirmed to contain the protein were collected and pooled. One volume of PreScission protease was added, and the solution was transferred to a dialysis membrane for
overnight exchange into a buffer without imidazole. The sample was then loaded back onto the 1 or 5 mL metal affinity column and processed as previously described. Without the histidine tag, the protein of interest will not bind to the metal ions and will be found in the flowthrough. The Precission Protease has its own histidine tag, causing it to bind to the resin and be separated from the truncated protein.

Ion exchange chromatography was also used on occasion. After metal affinity purification, the protein sample was dialyzed overnight into 20 mM Tris pH 8.5, 25 mM NaCl, and 1 mM DTT. The sample was loaded onto a Source-Q anion exchange column from GE Healthcare. A linear gradient of high salt elution buffer (20 mM Tris pH 8.5, 1 M NaCl, 1 mM DTT) was used over approximately 15-20 column volumes to disrupt ionic interactions and elute the bound protein.

Furthermore, size exclusion chromatography was used for a final polishing step. After completing at least one previous purification method, the sample was loaded onto a HiLoad™ 16/600 Superdex™ 200 column. A buffer of 20 mM Tris pH 8.5, 100 mM NaCl, and 1 mM TCEP was used for this method.

After each purification step, samples from the various fractions of the procedure were then loaded onto 12% or 15% SDS-PAGE gel to assess the purity of the target protein as previously described. The desired product was a clear, single band corresponding to the appropriate molecular mass for the target protein.

The concentration of a protein sample was determined using UV absorption. The Beer-Lambert law relates absorption (A) and concentration (c), as seen in Equation 2.1.

Equation 2.1: \[ A = bc \varepsilon \]
Aromatic residues such as tyrosine and tryptophan absorb at 280 nm\textsuperscript{20}, while nucleic acids absorb at 260 nm\textsuperscript{21}. Because the number of these residues can vary wildly between protein species, the extinction coefficient (\(\epsilon\)) is necessary to correct for such factors. The online resource ProtParam from ExPASy\textsuperscript{22} was used to obtain the appropriate extinction coefficient for each protein. The absorbance of each sample at 260 and 280 nm was measured using a Synergy H4 Hybrid Reader from BioTek. The 260/280 ratios were also noted. A ratio of 1.0 or less indicates that the sample is primarily composed of protein, while a ratio of greater than 2.0 indicates that the majority of the sample is composed of nucleic acids.

2.2 Mass spectrometry analyses

While SDS-PAGE can be employed to assess the presence and purity of a particular size of protein, it cannot be used to determine the protein’s identity. For this reason, mass spectrometry is often employed for identity confirmation.

2.2.1 MALDI: peptide fingerprinting

In this technique, a band in an SDS-PAGE gel is excised, and the protein extracted and digested. The sample is then analyzed by matrix associated laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. First, the band is excised with a razor blade, placed into a clean microcentrifuge tube, and washes of 500 \(\mu\)L 100 mM ammonium bicarbonate alternated with 500 \(\mu\)L acetonitrile are performed at room temperature, with rotating, to shrink the gel slice. The gel slice is placed into an Eppendorf Vacufuge Plus for 45 minutes until fully dehydrated. To digest the protein, 50 \(\mu\)L commercial-grade Trypsin in a 100 \(\mu\)M ammonium bicarbonate buffer is added, and
the sample is incubated at 37 °C for one hour. Once the digest is complete, the liquid is discarded and 50 mM ammonium bicarbonate is added to cover the gel slice. The sample is allowed to rotate overnight. The next day, the liquid is again discarded and 50 µL 5% acetonitrile and 0.1% TFA is added to extract the digested protein from the gel slice. After repeating this wash, 50 µL 50 acetonitrile and 0.1% TFA is added. The supernatants were saved, pooled, and dried in the Vacufuge until 10 µL remain. This sample was then analyzed using MALDI-TOF/MS. Using the sandwich method, alternating layers of α-cyano-4-hydroxycinnamic acid (CHCA) from Thermo Scientific and the sample were spotted onto a steel target plate. The plate is loaded into an UltrafleXtreme MALDI-MS from Bruker and the spectrum is observed and recorded. The fragments can then be compiled and a MASCOT analysis will compare the amino acid sequences of the sample to a database of known proteins23.

2.2.2 ESI

Electrospray ionization MS (ESI-MS) was utilized to observe product formation. For the enzymatic reaction, 260 nM MshC, 5 units Thermostatic Inorganic Pyrophosphatase from New England BioLabs, 200 µM ATP, 200 µM L-cysteine, and 100 µM GlcN-Ins were combined in a buffer of 25 mM imidazole, 50 mM NaCl, 5 mM magnesium sulfate, and 2 mM DTT and incubated at 37 °C for one hour. A control reaction omitting the GlcN-Ins was also prepared. To prepare the sample for mass spectrometry, 10 µL of reaction mix was quenched with 50:50 v/v of methanol and water to a total volume of 100 µL. The sample was injected onto an Esquire-LC/MS ion trap plus Hewlett Packard 1100 liquid chromatograph system for analysis. The resulting spectra were observed and recorded.
2.3 Activity assays

In order to research new drug molecules, one must first confirm the activity of the target enzyme. Once in vitro activity can be observed, enzyme kinetics can be determined to provide a baseline for future high-throughput screenings of potential inhibitors.

2.3.1 Malachite Green assay

The malachite green (MG) assay is an endpoint assay that measures the release of phosphate. In solution, the yellow MG absorbs at 446 nm. In the presence of acid and phosphate, MG forms a MG-phosphomolybdate complex that absorbs at 640 nm.

\[
P_1 + (\text{NH}_4)_2\text{MoO}_4 \xrightarrow{\text{H}^+} \text{H}_3\text{PMo}_{12}\text{O}_{40}
\]

\[
\text{H}_3\text{PMo}_{12}\text{O}_{40} + \text{HMG}^{2+} \xrightarrow{\text{H}^+} (\text{MG}^+)\left(\text{H}_2\text{PMo}_{12}\text{O}_{40}\right) + 2\text{H}^+
\]

yellow \(\rightarrow\) yellow, \(\lambda_{\text{max}} = 446\) nm

green \(\rightarrow\) green, \(\lambda_{\text{max}} = 640\) nm

Figure 2.2 Malachite green assay, A Malachite green salt, 4-[(dimethylamino)phenyl](phenyl)methylidene-N,N-dimethylcyclohexa-2,5-dien-1-iminium chloride. B MG reaction. Inorganic phosphate binds to ammonium molybdate in acidic conditions. Yellow malachite green binds with the yellow phosphomolybdate complex to form the green chromophore that absorbs at 640 nm.

For this assay, 2% malachite green solution and 6% w/v ammonium molybdate in 6 M HCl were combined in a 5.4:4.6 ratio. In the wells of a Greiner Bio-One 96-well U-shape microplate, 10 µM PhoRin in 20 mM HEPES pH 7.5 was mixed with 1.0 µM ATP.
to initiate the reaction. A negative control omitting the PhoRin was included. The well plate incubated at 37 °C for one hour, and was then quenched with 80 µL of the MG solution. After sitting at room temperature for 15 minutes to ensure complete quenching, the well plate was loaded onto the Synergy H4 Hybrid Reader and the absorbances were recorded.

To determine the concentration of released phosphate, a standard curve must be developed. A series of phosphate standards undergo the MG reaction, and the absorbances are plotted. By interpolation of the resultant graph, the phosphate concentrations of unknown samples can be obtained. A phosphate standard curve was determined using 2.0, 1.0, 0.50, 0.25, 0.125, and 0.0 µM solutions of sodium phosphate in 20 mM HEPES pH 7.5. For the reactions, 18 µL 20 mM HEPES pH 7.5 phosphate was mixed with 2 µL of the phosphate dilutions and incubated at 37 °C for one hour. Then 80 µL MG solution was added to quench the reaction, and the absorbance was taken after quenching progressed for 15 minutes. Each reaction condition was performed in triplicate.

2.3.2 Quinaldine Red assay

Quinaldine Red (QR) was utilized in the development of a high-throughput assay focusing on MshC drug development. The QR assay is an endpoint assay. The ligation catalyzed by MshC requires the hydrolysis of ATP to AMP and pyrophosphate. QR-molybdate binds phosphate\(^{24}\), requiring the coupling of MshC and inorganic pyrophosphatase to convert the pyrophosphate into phosphate. The reaction occurs in a Corning 384-well opaque white assay plate. This plate absorbs at 430 nm and emits at 530 nm\(^{24}\). The QR-phosphomolybdate chromophore absorbs at 530 nm, quenching the
emitted light of the plate\textsuperscript{24}. Thus, MshC activity can be detected by measuring the extent of the quenching.

The QR solution was composed of 0.05\% QR, 2\% w/v polyvinyl alcohol, 6\% w/v ammonium heptamolybdate tetrahydrate in 6 M HCl and water in a 2:1:1:2 ratio. The reaction mixture was composed of 260 nM \textit{M. smegmatis} MshC, 5 units of pyrophosphatase, 200 µM L-cysteine, and 200 µM GlcN-Ins in a buffer of 25 mM imidazole, 50 mM NaCl, 5 mM magnesium sulfate, and 2 mM DTT. A negative control omitting the GlcN-Ins and a second negative control omitting both GlcN-Ins and cysteine were also prepared. Finally, 200 µM of the potential inhibitor glucosazide inositol (GlcAz-Ins) was used in a fourth trial to replace the GlcN-Ins.

![Figure 2.3](image.jpg)

\textit{Figure 2.3} A potential inhibitor of MshC, glucosazide inositol

To initiate the reaction, 200 µM ATP was added to the well. At time points of 0, 5, 10, 15, 20, and 30 minutes, 40 µL of QR solution was added to each well and the well plate sat at room temperature for 2 minutes. Then, the reactions were quenched by the addition of 2 µL 32\% sodium citrate. The well plate was incubated at 37 °C with light agitation for 15 minutes, and then loaded onto the Synergy H4 Hybrid Reader. The relative fluorescent units were observed and recorded. A standard curve for phosphate was developed using sodium phosphate.
After the optimum reaction time was determined, the optimum enzyme concentration was assayed. The reaction mixture was composed of 5 units of pyrophosphatase, 200 µM L-cysteine, and 200 µM GlcN-Ins in a buffer of 25 mM imidazole, 50 mM NaCl, 5 mM magnesium sulfate, and 2 mM DTT. *M. smegmatis* MshC
concentration was varied from 0 nM to 500 nM. The reaction was initiated with 200 µM ATP and incubated at room temperature for 30 minutes before quenching as described previously. The white plate was loaded onto the Synergy H4 Hybrid Reader and the relative fluorescent units were observed and recorded.

2.3.3 Fluorescent Polarization assay

This assay was intended to assess the in vitro activity of PhoP. Because PhoP is a DNA binding protein, 200 nM fluorescein labeled stem-loop DNA was combined with PhoP ranging from 0.1 µM to 45 µM in 20 mM Tris pH 8.5, 100 mM NaCl, 25 mM imidazole, and 5 mM BME. The DNA probe had a sequence of 5’–CCTAATTATAACGAAGTTATAATTAGG – 3’. The reactions were incubated at room temperature for 15 minutes before and were then loaded into the Hybrid Reader. The fluorescent polarization was measured and recorded at an excitation wavelength of 485 nm and an emission wavelength of 528 nm.

2.3.4 Pull-Down assay

For this assay, 1 L cultures of PhoRin and PhoP were grown, harvested and prepared for purification using the Batch Method as described in Section 2.1.2. PhoRin was dialyzed into the binding buffer (20 mM Tris pH 8.5, 25 mM imidazole, 100 mM NaCl, 5 mM BME). PhoP was combined with 1x Precission Protease and also dialyzed into the equilibration buffer. After dialysis, the PreScission protease and histidine tag were separated from PhoP using the batch method. To run the pull down, 20 µM of both PhoRin and PhoP were combined with 20 µM AMP-PNP and incubated at room temperature for 30 minutes. A control consisting of both enzymes with no substrate was included. The samples were loaded onto nickel resin and purified using the batch method.
The reserved flowthrough and eluted fractions were loaded onto a 15% SDS-PAGE gel for observation.

2.4 Crystallization strategy

Crystallization attempts were performed using the hanging-drop method. Proteins used in crystallization were between 0.1 and 20 mg/mL in concentration (with a median of 5 mg/mL) and were dialyzed into 20 mM Tris pH 8.5, 100 mM NaCl, 25 mM imidazole, and 5 mM BME prior to placement in the crystal tray. Conditions for each reservoir of the well plates were obtained from the Hampton Research HR2-144 Index Screen. Preliminary crystals were optimized using 1-dimensional screens: the precipitant present in the well solution was varied from 5 to 95% of its concentration in the original condition. The Hampton Research HR2-428 Additive Screen was also used to optimize crystallization conditions.

Substrates and substrate analogues were also employed to encourage crystallization. These additions will be discussed thoroughly in subsequent chapters. Crystal trays were stored at approximately 20 °C.

For X-ray diffraction experiments, all crystals were cryoprotected by the addition of 50% glycerol to the crystallization drop if the well solution did not contain a suitable cryoprotectant. The crystals were then flash-cooled with liquid nitrogen and shipped to Argonne National Laboratories for remote analysis using the Advanced Photon Source.
Chapter 3

PhoPR Results

3.1 Expression and purification of PhoP and PhoR

*PhoP*, *phoR*ex, and *phoR*in were amplified from H37Rv using traditional PCR and were each inserted into pET28 vectors. PhoR was separated into PhoRex and PhoRin based upon secondary structure prediction software\textsuperscript{25,26}. The recombinant DNA was expressed using Rosetta cells. All cells were grown in Luria-Bertani broth supplemented with kanamycin and chloramphenicol.

Table 3.1 Sequences of forward and reverse primers for *phoP*, *phoR*ex, and *phoR*in. *PhoP* and *phoR*ex primers include restriction sites for NdeI and HindIII; *PhoR*in primers include restriction sites for NdeI and BamHI-HF.

<table>
<thead>
<tr>
<th></th>
<th><em>phoP</em></th>
<th><em>phoR</em>ex</th>
<th><em>phoR</em>in</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-</td>
<td>CAC CCC AGC ATA TGC GGA AAG GGG TTG ATC TCG - 3’</td>
<td>CAC CTT CAT ATG TTG CAG CAC CGG CTG ACC A - 3’</td>
<td>CAC CTT CAT ATG CGC CGC AGC CTG CGG CCG CT - 3’</td>
</tr>
<tr>
<td>5’-</td>
<td>CCA GAA GCT TTC ATC GAG GCT CCC GCA GTA C - 3’</td>
<td>TTT AAG CTT TCA TGA CCG CAC GGT GCT CCG GA - 3’</td>
<td>TTG GAT CCT CAG GGC GGC CCT GGC ACA ACT GGC GT - 3’</td>
</tr>
</tbody>
</table>

All three proteins were cloned to include a poly-histidine tail at the N-terminus, allowing for binding to a metal affinity column. Each protein was purified with nickel affinity chromatography. While monitoring A\textsubscript{280}, a single, large peak was observed in the eluted fraction starting around 20% elution buffer. After a SDS-PAGE gel confirmed the
presence of the desired protein, the corresponding fractions were collected and dialyzed in the presence of PreScission protease as described previously. When histidine tag removal was performed, the flowthrough fraction was confirmed to be the desired protein without the tag. PhoRin was purified further with size exclusion chromatography, and the fractions were analyzed with SDS-PAGE to confirm the presence of PhoRin.

Figure 3.1 Example of a PhoP nickel affinity chromatogram featuring a step gradient; $A_{280}$ is in blue, salt concentration is in tan, and conductivity is in green.
While spectrophotometrically determining the concentration of PhoRex after metal affinity chromatography, a heightened 260/280 ratio of approximately 1.5 was observed. The excess nucleic acids were removed from the sample by applying a high salt wash (20 mM Tris pH 8.5, 500 mM NaCl, 25 mM imidazole, and 5 mM BME) to the nickel column after PhoRex was loaded. Protein elution with the usual high imidazole buffer followed.

3.2 MALDI peptide fingerprinting

The expression of PhoP, PhoRex, and PhoRin were confirmed by MALDI-TOF/MS peptide fingerprinting. MASCOT results indicated approximately 33% sequence coverage for PhoRex and 60% for PhoRin.
### 3.3 Malachite green activity assays

PhoR autophosphorylates a conserved histidine residue. The malachite green (MG) assay is a useful technique for determining the spontaneity of histidine dephosphorylation. Triplicate reactions using 10 µM PhoRin were run along side the negative control. The negative control included all reaction components except PhoRin itself. After 1 hour incubation at 37 °C, the reactions were quenched with the MG solution. Table 3.2 depicts the obtained absorbances.

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>+PhoRin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absorbance</td>
<td>0.169 ± 0.0042</td>
<td>0.643 ± 0.036</td>
</tr>
</tbody>
</table>

The average absorbance of the reaction samples was 0.620 units with a standard deviation of 0.036. From the phosphate standard curve, the concentration of released phosphate was calculated as approximately 1.1 µM. This indicates a reaction stoichiometry of approximately one turnover per ten enzyme active sites.

![Phosphate standard curve](image)

**Figure 3.3** Phosphate standard curve
3.4 Fluorescence polarization activity assay

The polarization readings of this assay are described in Table 3.3. PhoP in concentrations of 0.0 to 45 µM was combined with 200 nM fluorescein labeled stem-loop DNA and incubated for 15 minutes.

Table 3.3 Results of PhoP-DNA fluorescence polarization assay

<table>
<thead>
<tr>
<th>[PhoP], uM</th>
<th>0.0</th>
<th>0.1</th>
<th>0.5</th>
<th>1.0</th>
<th>5.0</th>
<th>10</th>
<th>20</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polarization, mP</td>
<td>162</td>
<td>168</td>
<td>154</td>
<td>167</td>
<td>164</td>
<td>195</td>
<td>182</td>
<td>191</td>
</tr>
</tbody>
</table>

These data indicate that the PhoP-DNA complex is not forming, as we do not observe a concentration-dependent change in fluorescence polarization.

3.5 PhoPR complex

PhoRin and PhoP were purified using nickel affinity resin. The purification of each protein was confirmed via SDS-PAGE. The histidine tag of PhoP was removed, and the presence of PhoP in the flowthrough fraction was confirmed via SDS-PAGE as well. Both proteins were concentrated to 20 µM. A control of PhoRin plus PhoP in binding buffer was mixed with the nickel resin, while the reaction solution of PhoRin, PhoP, and AMP-PNP was mixed with a separate nickel resin. After batch method purification, the flowthrough and eluted fractions of each trial were assessed with SDS-PAGE. PhoP was observed in the flowthrough, and PhoRin was observed in the eluate for both the control and the reaction. This indicates the absence of a stable PhoRin-PhoP complex under these conditions.
3.6 Crystallization attempts

PhoRex crystals formed in Index Screen Conditions 1, 15, 16, 55, and 85. Except for Condition 1, all PhoRex sample used in crystallization attempts underwent a high salt wash during metal affinity chromatography and poly-histidine tag removal.

A small, 30 µm isometric crystal formed in Condition 1 (0.1 M citric acid pH 3.5, 2.0 M ammonium sulfate) at a PhoRex concentration of approximately 7 mg/mL. The protein sample used had an intact poly-histidine tag and did not experience a high salt wash during purification. This crystal required seven months to grow, and did not diffract well. To optimize growth conditions, a one-dimensional screen was established to vary the ammonium sulfate concentration in Condition 1 from 0.05 M to 2.0 M. No further crystallization was observed in this screen.

Condition 15 (0.1 M HEPES pH 7.5, 0.5 M magnesium formate dihydrate) yielded octahedral crystals approximately 40 µm in length and 10 µm at the widest point. The concentration of PhoRex for this condition was approximately 20 mg/mL.

Figure 3.4   SDS-PAGE gel of the pull down method. Both the negative control and the reaction indicated PhoP in the flowthrough and PhoRin in the eluate.
Precipitation was also observed, and the crystals required six months to grow. A one-dimensional screen varying the concentration of the well solution from 10 to 100% was set up. Crystals of similar size and morphology to those of the original condition formed in each dilution, with the cleanest and most numerous in 30% and 40% fractions. However, the precipitation did not improve, and the crystals still needed six months to develop. These crystals diffracted poorly and no useable data was obtained.

Condition 16 (0.1 M Tris pH 8.5, 0.3 M magnesium formate dihydrate) yielded rhombic plates of approximately 100 µm long and 20 µm wide. PhoRex was approximately 20 mg/mL in this condition, and a moderate amount of precipitation was observed. These crystals required 6 months to grow. A one-dimensional screen varied the concentration of the well solution from 10 to 100%. Crystals of similar size and morphology to those of the original condition formed in each dilution, with the cleanest and most numerous in the 30% fractions. Precipitation was lessened with dilution, although it did not disappear entirely. The crystals still needed six months to develop. These crystals diffracted poorly and no useable data was obtained.

After one year of incubation, Condition 55 (0.2 M magnesium chloride hexahydrate, 0.1 M HEPES pH 7.5, 30% w/v PEG monomethyl ether 550) yielded a single, tetrahedral crystal of approximately 300 µm long and 30 µm wide. This crystal features a large imperfection in its lower half. PhoRex was concentrated to 13 mg/mL for this condition, and much precipitation is observed. This crystal diffracted poorly and no useable data was obtained.

Condition 85 (0.2 M magnesium chloride hexahydrate, 0.1 M Tris pH 8.5, 25% w/v PEG 3350) appears to be very similar to conditions 16 and 55. Several tetrahedral
crystals of approximately 200 µm in length and 30 µm in width are present, along with some clusters of 200 µm long needles. Some precipitation and aggregation is apparent, though to a lesser extent than in conditions 15 and 55. This condition also required one year for crystal growth to become prominent. These crystals diffracted poorly and no usable data was obtained.

Figure 3.5  Samples of PhoRex and PhoRin crystals
   A) PhoRex in 0.1 M citric acid pH 3.5, 2.0 M ammonium sulfate (Condition 1)
   B) PhoRex in 30% 0.1 M HEPES pH 7.5, 0.5 M magnesium formate dihydrate (Condition 15)
   C) PhoRex in 30% 0.1 M Tris pH 8.5, 0.3 M magnesium formate dihydrate (Condition 16)
   D) PhoRex in 0.2 M magnesium chloride hexahydrate, 0.1 M HEPES pH 7.5, 30% w/v PEG monomethyl ether 550 (Condition 55)
   E) PhoRex in 0.2 M magnesium chloride hexahydrate, 0.1 M Tris pH 8.5, 25% w/v PEG 3350 (Condition 85)
PhoRin was purified on the size exclusion column and was easily concentrated up to and beyond 20 mg/mL. Heavy precipitation was frequently observed, even at lower concentrations of PhoRin. A series of index screens including either 0.2 mM ATP or 0.2 mM AMP-PNP were performed, however, crystallization has not yet been observed.

3.7 Discussion

Secondary structure predictions of PhoR indicated the location of the transmembrane domain. This domain, consisting of three α-helices, was omitted from the construct due to its inherent incompatibility with the extracellular and intracellular domains in vitro. The extracellular and intracellular domains are hydrophilic, while the transmembrane domain is hydrophobic. Without the plasma membrane, it is extremely difficult to ensure that the protein would fold correctly to produce a stable, functional enzyme. Typically, yeast, insect, or mammalian cell lines are required for the expression of transmembrane proteins27. As our laboratories were not equipped to support these cell lines, it was not practical to work with the intact PhoR.

PhoP, PhoRex and PhoRin are readily expressed in E. coli hosts and easily purified using metal affinity and size exclusion chromatography. The 260/280 ratio of 1.5 observed after nickel affinity purification of PhoRex suggests the presence of significant amounts of nucleic acid. It was thought that confirming the identity of said nucleic acid could aid in the determination of the PhoRex substrate(s). MALDI-TOF/MS performed on a sample of PhoRex identified a peak at 442.723 m/z, coinciding with the molecular mass of GDP (443.01 amu). The observation that PhoRex may be recognizing GDP was quite interesting, as GDP is involved in an important step of phagosome maturation28. One of the mechanisms of survival for the Mycobacterium tuberculosis bacterium is
prevention of phagosome maturation. Unfortunately, further mass spectrometry analyses were unable to confirm the presence of GDP or any other specific nucleotide or nucleic acid. Therefore, the high salt wash step was developed to remove the contaminants and facilitate future crystallization attempts.

The MG assay suggests that PhoRin is dephosphorylated spontaneously, with one turnover per ten enzyme active sites. Approximately 1.1 µM of phosphate was released per 10 µM PhoRin over 30 minute reactions. This indicates a low activity of PhoRin in vitro. Because PhoRin has been truncated to omit the PhoR extracellular and transmembrane domains, it is probable that its ATPase activity is not significant without reception of an environmental stimulus or the ability to transfer phosphate to PhoP.

No PhoP activity was observed in the FP assay. A variety of factors could account for the lack of activity. It is possible that PhoP was not folding correctly during expression and the DNA binding site was therefore unable to interact with the DNA. However, the simplest and most plausible explanation is that the sequence chosen for the binding DNA sample did not accurately represent the promoter(s) that PhoP recognizes. As mentioned previously, the method in which PhoP binds DNA is currently unknown. The sequence(s) that PhoP recognizes is also unknown. However, recently, Galagan and coworkers suggested a DNA binding motif for PhoP using chromatin immunoprecipitation followed by sequencing (ChIP-Seq)\textsuperscript{29}. To briefly describe this procedure, an antibody specific to the targeted DNA binding protein is used to precipitate said protein\textsuperscript{30}. The DNA bound to the protein at the time of immunoprecipitation is then isolated for sequence analysis\textsuperscript{30}. Using this method, the probable DNA binding motif of
PhoP was reported to be 5’– CTGTGAAXTXGCTGXXXAX – 3’, where X indicates a nonspecific nucleobase\textsuperscript{30}. A new DNA probe will be designed according to this data.

The pull-down assay was developed to take advantage of affinity chromatography in order to observe protein-protein interactions. PhoRin is expected to interact with PhoP through phosphate transfer. During purification, one enzyme will undergo cleavage of its histidine tag, while the other enzyme’s tag will remain intact. Adenosine 5’-(β,γ-imido)triphosphate (AMP-PNP), a non-hydrolysable ATP analogue, was utilized with the goal of trapping the PhoRin-phosphate-PhoP transition state. The adenosine moiety should remain bound in the PhoRin active site, while the nonhydrolyzable phosphate is captured by PhoP, resulting in the tethering of the two enzymes. The sample can then be loaded onto an affinity resin: if the enzyme complex was established, both enzymes should be observable in the eluted fraction. However, if the tag-less enzyme appears in the flowthrough, then a stable complex did not form.

In the pull-down experiment, PhoP was observed in the flowthrough and PhoRin was observed in the eluate. This indicates that the use of AMP-PNP was not sufficient to hold a PhoPR complex together. The catalytic aspartate of PhoP may be blocked from accepting the phosphate transfer due to the geometry of the PhoRin active site while AMP-PNP is bound. It is also worth noting that \textit{in vitro} activity of PhoP was not observable in other experiments; that PhoP is not folding properly \textit{in vitro} is another possible explanation of the lack of an observable complex. Finally, some research suggests that AMP-PNP and other non-hydrolysable analogs may act as competitive inhibitors of two component systems\textsuperscript{31}, and therefore by their inherent nature prevent the complex between the HK and RR from forming. The HK-RR phosphotransfer transition
state is transient and does not experience tight binding; attempts to capture this fleeting complex may be fundamentally futile.

Crystallization of PhoRex and PhoRin requires a significant amount of optimization. Currently, PhoRex crystals require approximately 6-12 months to form, while crystallization of PhoRin is not yet successful. Preliminary PhoRex crystals diffract poorly and occasionally present with hollow impurities, indicating rapid crystal growth. This suggests that nucleation is the rate-limiting step in these attempts and is the causative agent in the excessive length of time required for the development of these crystals. PhoRex appears to favor crystallization conditions that include magnesium or a low pH. Magnesium is a known substrate of some PhoR homologs\(^12\). It is possible that PhoR also recognizes magnesium, however we will be unable to confirm this proposal without a solved crystal structure. Considering the domain’s exposure to the acidic environment of the phagosome, another possibility is that PhoR recognizes pH change. A sudden drop from physiological pH would indicate assimilation into the phagosome, leading to the initiation of \textit{M.tb} survival mechanisms. Indeed, some histidine kinases are known to recognize pH changes. Solving the crystal structure of PhoR will elucidate much about its substrate(s) and subsequent mechanisms.
Chapter 4

MshC Results

4.1 Expression and purification of MshC

The *M. smegmatis* *cysS* was synthetically derived and amplified using traditional PCR followed by insertion into pET28. The recombinant DNA was expressed using T7 Rosetta cells grown in Luria-Bertani broth supplemented with kanamycin and chloramphenicol.

Both genes were cloned to include a poly-histidine tag at the N-terminus to allow for metal affinity chromatography. *M. smegmatis* MshC was purified by cobalt affinity followed by anion exchange chromatography and finally size exclusion chromatography. Protein purity was monitored through SDS-PAGE at each purification step. Generous

<table>
<thead>
<tr>
<th><strong>M. smegmatis cysS</strong></th>
<th>5’– CAC CAT ATG CAA TCG TGG TCG GCA – 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5’– CTC GAG TTA GAG GTC CAC ACC CAG CA – 3’</td>
</tr>
<tr>
<td><strong>M. tb mshC</strong></td>
<td>5’– CTG TTC CAG GGA CCT CAG AGC TGG TAT TGT CCG CCT – 3’</td>
</tr>
<tr>
<td></td>
<td>5’– CCC TGC TGG GTG TTG ATC TGT AAG CGT CCG GAT CCG AA – 3’</td>
</tr>
</tbody>
</table>
yields of MshC were collected and found to be stable at 4 °C for approximately four months before the onset of degradation.

Figure 4.1  Example of an MshC cobalt affinity chromatogram. MshC was loaded and purified on a 1 mL cobalt column using a linear gradient.
M. tb mshC was synthetically derived and amplified using traditional PCR. The gene was inserted into pET32 tagged with thioredoxin (TRX), pET42 tagged with glutathione S-transferase (GST), and pET28 tagged with maltose binding protein (MBP) using the Gibson assembly method. BL21* and Rosetta cells were transformed with the Gibson construct. Currently, no expression has been achieved.

4.2 Confirmation of product formation using ESI-MS

The reaction of M. smegmatis MshC with its substrate was analyzed using ESI-TOF/MS. Cys-GlcN-Ins has a molar mass of approximately 444 amu. A peak at 445 m/z was observed in the reaction spectra, while no such peak was observed in the chromatogram of the control. This corresponds to the protonated mass of Cys-GlcN-Ins.
The QR assay was optimized for conditions of 260 nM MshC, 5 units of pyrophosphatase, 200 µM L-cysteine, and 200 µM GlcN-Ins in a buffer of 25 mM imidazole, 50 mM NaCl, 5 mM magnesium sulfate, and 2 mM DTT. The reactions were initiated with 200 µM ATP. The substrate concentrations were chosen based on published $K_m$ values.$^{17}$ A trial including 200 µM of the potential inhibitor glucosazide inositol (GlcAz-Ins) in the place of GlcN-Ins was included. Two negative controls – one omitting GlcN-Ins and a second omitting both GlcN-Ins and Cys – were also performed.

The reaction of MshC was observed to experience approximately 85% quenching of signal at the end of the thirty minutes. The negative controls were not active. The sample using the GlcAz exhibited only 20% quenching.

Figure 4.3 ESI spectra indicating the protonated mass of Cys-GlcN-Ins at approximately 445 m/z
In the second QR activity assay, MshC concentration was varied while holding the reaction time constant at 30 minutes. As the concentration of enzyme increased, the degree of quenching also increased. A maximal reaction rate was reached at approximately 300 nM MshC. For subsequent assays, a concentration of 250 nM MshC was chosen.

![Graph](image)

**Figure 4.4** QR MshC assay including complete enzymatic reaction (diamonds), negative control omitting GlcN-Ins (squares), negative control omitting GlcN-Ins and cysteine (triangles), and enzymatic reaction with GlcAz-Ins (circles).

In the second QR activity assay, MshC concentration was varied while holding the reaction time constant at 30 minutes. As the concentration of enzyme increased, the degree of quenching also increased. A maximal reaction rate was reached at approximately 300 nM MshC. For subsequent assays, a concentration of 250 nM MshC was chosen.

![Graph](image)

**Figure 4.5** QR assay varying the concentration of MshC for a reaction time of 30 minutes: Quenching was monitored and indicated saturation at approximately 300 nM MshC.
4.4 Crystallization attempts

*M. smegmatis* MshC was co-crystallized with 2 mM ATP, 2 mM L-cysteine, and 0.2 mM GlcAz-Ins using Index Screen conditions. Protein concentrations varied from 5 mg/mL to 20 mg/mL. Thus far, no crystals have formed.

4.5 Discussion

4.5.1 Discussion of results

*M. smegmatis* MshC expresses readily and liberally. Protein after size exclusion chromatography is very pure and stores at 4 °C for at least four months. When combined with 50% glycerol, MshC remained viable for over six months at -80 °C. The generous yields of MshC plus its longevity are an appreciable convenience for working with this protein.

*M.tb* MshC has proven to be a true challenge. The expression of soluble protein has evaded our reach at the preparation of this thesis. Though our molecular cloning utilized codon-optimized Rosetta cells for expression, recent studies suggest *M.tb* MshC

---

Figure 4.6 Phosphate standard curve for the QR assay

![Phosphate standard curve](image)
cannot be heterologously expressed in *E. coli* cells\(^2\), regardless. Besides codon bias, other factors impacting heterologous expression may include degradation of the foreign protein by host proteases or the formation of inclusion bodies. However, other groups such as Gutierrez-Lugo *et al.* and Newton *et al.* have developed a construct of MshC that does express successfully\(^3\). For this construct, *M.tb* MshC was expressed in a fusion construct with MBP in the I64 strain of *Mycobacterium smegmatis*\(^3\). This *M. smegmatis* strain has the point mutation L205P in the *mshC* gene, rendering it unable to produce mycothiol\(^5\). Therefore, expression using I64 *M. smegmatis* proves advantageous.

Using ESI-MS, we were able to confirm the *in vitro* activity of *M. smegmatis* MshC. The Cys-GlcN-Ins product was observed, indicating that this enzyme is folded correctly and able to synthesize its product.

MshC releases one molecule of pyrophosphate per one molecule of Cys-GlcN-Ins, and inorganic pyrophosphatase converts one molecule of pyrophosphate into two molecules of phosphate. In the QR assay, one QR-phosphomolybdate complex forms for each phosphate. Therefore, the rate of Cys-GlcN-Ins production is proportional to the rate of QR-phosphomolybdate quenching in a 1:2 ratio. When the concentration of MshC was varied while maintaining a 30 minute reaction time, a maximal reaction rate was observed at approximately 300 nM MshC. This indicates that at 300 nM MshC, the rate-limiting step of the reaction is catalyzed by the pyrophosphatase. Because the assay is only useful when the rate-limiting step is catalyzed by our enzyme of interest (MshC), an MshC concentration of 250 nM will be utilized in subsequent assays. These data will act as a baseline for testing the inhibition properties of potential drug molecules. The potential inhibitor, GlcAz-Ins, was demonstrated to impede product formation, though
some MshC activity was still observed. This can be attributed to the reduction of the azide by DTT.

The crystal structure of *M. smegmatis* MshC, as solved by Tremblay and coworkers, was obtained from a truncated enzyme\(^\text{19}\). A flexible loop in the stem contact fold (residues 285-297) was proteolyzed\(^\text{19}\). Because this region is predicted to have significance in adenylate formation and GlcN-Ins orientation\(^\text{19}\), it is desirable to solve an intact MshC crystal structure that includes this domain. For this reason, MshC was co-crystallized with ATP, cysteine, and GlcAz-Ins. The inhibitor was intended to stabilize the loop of the SC fold by trapping MshC in a Cysadenylate/GlcN-Ins ternary complex. While GlcAz-Ins appears to inhibit MshC activity (see MG assay), its ability to trap the MshC ternary complex is unconfirmed.

### 4.5.2 Current enzyme kinetics

In 2007, Fan and coworkers published the steady state and pre-steady state kinetics of *M. smegmatis* MshC\(^\text{17}\). AMP formation was coupled with myokinase, pyruvate kinase, and lactate dehydrogenase to determine the initial velocities of MshC\(^\text{17}\). Myokinase transfers a phosphate to AMP, forming ADP\(^\text{36}\). Pyruvate kinase then transfers a phosphate from phosphoenolpyruvte to ADP to produce ATP and pyruvate\(^\text{37}\). Finally, lactate dehydrogenase converts pyruvate to lactate via the dehydration of NADH to NAD\(^+\)\(^\text{38}\). Two molecules of NADH are oxidized for each molecule of Cys-GlcN-Ins produced by MshC, thus by monitoring the decrease in absorbance of NADH at A\(_{340}\), MshC kinetic data can be obtained\(^\text{17}\). The reaction conditions involved 100 mM HEPES pH 7.8, 10 mM MgCl\(_2\), 10 mM L-cysteine, 2 mM DTT, 100 μM GlcN-Ins, 1 mM
potassium phosphoenolpyruvte, 200 µM NADH, 18 units myokinase, 18 units pyruvate kinase, and 18 units lactate dehydrogenase. The addition of 25 nM MshC initiates the reaction and the absorbances were measured. To determine the initial velocities, the concentration of one substrate was varied while keeping fixed concentrations of a second substrate and a saturating concentration of the third. Double-reciprocal plots of the initial velocities using ATP or cysteine both reveal parallel lines. This supports a Ping Pong mechanism with the formation of a ternary MshC-ATP-cysteine complex. The $K_m$ values of ATP, cysteine, and GlcN-Ins were determined to be approximately 1.8, 0.1, and 0.16 mM, respectively.

In one inhibition assay, pyrophosphate was used with varying concentrations of ATP, 500 µM GlcN-Ins, and either 40 µM or 2 mM cysteine. In the trial of pyrophosphate versus ATP under a non-saturating concentration of cysteine (40 µM), a non-competitive inhibition pattern was observed with a $K_{ii}$ value of 3.7 ± 1.0 mM and a $K_{is}$ value of 3.4 ± 1.0 mM. Conversely, when pyrophosphate inhibition was assayed versus ATP under a saturating cysteine concentration (2 mM), an uncompetitive inhibition pattern was observed with a $K_{ii}$ value of 1.9 ± 0.1 mM. These data and observations support ATP binding to MshC first, with cysteine binding second.

In a second inhibition assay, CSA was analyzed as a potential bi-substrate inhibitor of MshC versus varying concentrations of either ATP (with 50 µM cysteine and 100 µM GlcN-Ins) or cysteine (with 1 mM ATP and 100 µM GlcN-Ins). CSA was determined to act as a competitive inhibitor versus ATP, and a non-competitive inhibitor versus cysteine. The competitive inhibition constant was determined to be 304 ± 40 nM versus ATP, and the $K_{is}$ and $K_{ii}$ for inhibition versus cysteine were 2.4 ± 0.2 µM and 390
± 60 nM, respectively\textsuperscript{17}. This confirms the binding pattern of ATP and cysteine, and fully supports a Bi Uni Uni Bi Ping Pong mechanism, as detailed below in Figure 4.4.

![Figure 4.7](Image)

**Figure 4.7** Schematic of the MshC Bi Uni Uni Bi Ping Pong mechanism

Ping pong reactions are composed of two half reactions whose individual rates are greater than the rate of the steady state\textsuperscript{17}. Therefore, pre-steady state kinetics of *M. smegmatis* MshC were also determined. In the first half reaction, the rate of adenylate formation was measured by mixing MshC with 1 mM L-cysteine, 10 mM DTT, 200 mM HEPES pH 7.8, 0.1 to 4 mM radiolabeled ATP ([\(\alpha\)-\(33\)P]-ATP, 20 µCi/µmol), 10 mM MgCl\(_2\), and 100 mM HEPES pH 7.8\textsuperscript{17}. The reaction was quenched with 110 µL 150 mM EDTA and the enzyme denatured by boiling\textsuperscript{17}. TLC resolved the radiolabeled ATP and AMP using 0.9 M guanidine hydrochloride as a mobile phase\textsuperscript{17}. The rate of cysteine-adenylate formation was determined to be 9.4 ± 0.7 s\(^{-1}\), and the \(K_d\) for ATP was measured to be 1.7 ± 0.2 mM.

For the second half reaction, the rate of cysteine ligation was measured by mixing 51 µM MshC, 20 mM MgCl\(_2\), 2 mM ATP, 1 mM radiolabeled cysteine ([\(14\)C]-L-cysteine, 50 µCi/µmol), 10 mM DTT, and 100 mM HEPES pH 7.8\textsuperscript{17}. After incubation at 25 °C, the solution was combined with 40 to 1000 µM GlcN-Ins in 100 mM EDTA and 100 mM HEPES pH 7.8\textsuperscript{17}. The reaction was quenched with 110 µL 3% TFA, boiled to
denature the enzyme, and separated using TLC with a mobile phase of CH₃CN:H₂O:28% NH₃ in a 3:1:1 ratio. Two radiolabeled species were separated: [¹⁴C]-L-cysteine and [¹⁴C]-L-cysteine-GlcN-Ins. The rate for the second half reaction was determined to be 5.2 ± 0.8 s⁻¹ and the K_d for GlcN-Ins was determined to be 290 ± 90 uM.
Chapter 5

Conclusions and Future Work

The PhoPR two-component system is essential to the virulence of *Mycobacterium tuberculosis*. As this system is not present in eukaryotes, PhoP and PhoR present attractive targets for drug design. PhoR was expressed as separate extracellular and intracellular domains due to its inherent *in vitro* instability as a transmembrane protein. Diffracting protein crystals have not yet been obtained, though PhoRin is expected to have a structure similar to other cytoplasmic HK domains, including the conserved α-β sandwich the N, G1, and G2 boxes to form the ATP-binding domain. PhoRin has measurable autophosphorylation and dephosphorylation activity. PhoP was expressed as the native, full-length protein. Current crystallization data indicated that PhoP forms a head-to-head homodimer, while its mechanism for DNA binding remains uncertain. Fluorescence polarization assays did not indicate DNA-binding activity of PhoP, though this may be due in part to the non-specificity of the DNA probe. PhoRin and PhoP were not observed to form the phosphotransition state using the pull-down method plus non-hydrolyzable ATP analog, AMP-PNP. Currently, our crystallization attempts of PhoRex and PhoRin have proved unsuccessful.
We will continue our crystallization attempts in the hopes of obtaining PhoRex and PhoRin crystals that readily diffract. The crystallization of PhoRex remains our principal goal for the future; elucidating its active site and substrate(s) will allow for the development of high-throughput assays and the identification of potential drug molecules. Other nonhydrolyzable ATP analogs, including but not limited to, adenosine 5′-[γ-thio]triphosphate (ATP-γ-S) should be used for future crystallization attempts of PhoRin. ATP-γ-S may also be used as a basis for a second pull-down assay to evaluate the PhoRin-PhoP complex. If this complex can be observed and stabilized, its crystallization could prove significant in the advancement of our understanding of this TCS. The recent DNA binding motif for PhoP will also be utilized in the design of an optimized DNA probe. Inhibition of the PhoPR system will result in an avirulent bacterium with a weakened cellular envelope, increasing susceptibility to current antituberculin drugs.

Mycothiol biosynthesis is unique to actinomycetes and essential for their survival. Mycobacterial strains lacking MshC demonstrate hypersensitivity to antibiotics, as well as failure to thrive in the case of *M. tb*. Thus, this enzyme presents another attractive target for drug design. *M. smegmatis* MshC was expressed generously in its native form. MshC catalyzes the penultimate step in mycothiol biosynthesis via a Bi Uni Uni Bi Ping Pong kinetic mechanism. The reaction mechanism encompasses two half reactions: first, the formation of a cysadenylate intermediate, and second, the ligation of this intermediate to GlcN-Ins. Using the quinaldine red assay, we were able to demonstrate MshC’s activity *in vitro* as a scaffold for future high-throughput drug screening. While our
crystallization attempts were unsuccessful, a truncated *M. smegmatis* MshC crystal structure has been obtained elsewhere.

We will continue to develop the QR assay for high-throughput screening of clinical collections. Also, currently in motion is an alternate assay involving a fluorophore-bound cysteine allowing MshC activity to be monitored using FP. Briefly, cysteine is covalently modified with a fluorescent probe (Alexa Fluor® 488). Monitoring the FP will allow the activity of MshC to be tracked in the presence of potential inhibitors. If a tested drug is able to displace the probe, the polarization will decrease. Crystallization attempts will continue to develop as well. As potential inhibitors of MshC are identified, co-crystallization using these compounds will proceed to hopefully result in the refinement of the full-length MshC crystal structure.

The expression of *M. tb* MshC still remains our goal. While successfully expressed in I64 *M. smegmatis*, expression in *E. coli* competent cells will fit our needs more efficiently. The use of synthetic DNA, codon-optimized competent cells, and Gibson assembly will hopefully allow for the expression of soluble, stable protein. The inclusion of various tags such as MBP, TRX, and GST are intended to further aid in these attempts. Once *M.tb* MshC is expressed and purified, we can replicate our QR assay to determine enzyme kinetics, with the eventual goal of screening clinical collections for potential drug molecules.

Though our PhoPR and MshC projects remain in relative infancy, we have set up the foundations for significant innovation and drug development. PhoR, PhoP, and MshC individually present impressive drug targets in the ever-evolving fight against
tuberculosis. Development of inhibitors for any one of these three enzymes could be momentous.
References


32. Gutierrez-Lugo, M. T.; Newton, G. L.; Fahey, R. C.; Bewley, C. A., Cloning, expression and rapid purification of active recombinant mycothiol ligase as B1


Appendix A

phor and PhoR sequences

Gene sequence
PhoRex is highlighted in yellow, and PhoRin is highlighted in green

*M. tuberculosis* H37Rv|Rv0758

*phoR*: 1458 bp
**Protein sequence**
PhoRex is highlighted in yellow, and PhoRin is highlighted in green

*M. tuberculosis* H37Rv|Rv0758
PhoR: 485 aa

MARHLRGRPLRVLVATLLVATGLVASGIAVTSMLQHRLTSRIDRVLLEEAAQIWAQITLPLAPDPYPGHNPDRPPSRFYVRVISPDGQSYTALNDNTAIPAVPANNDVGRHPTTLPSSGSKTLWRAVSVRASDGYLTTVAILDADVSTVRSLVLLQVIGS

AVLVVPGVAGYAVVRRSLPLAEFETTAADAGAGQLDRVPQWHPRTVEGRLSLALNGMLAQIQRAVASAESSAEKARDSEDRMQFIIDASHELRTPLTTIIRGFAELYRQGAARDVGMLLSRCIESASRMGLLVDELDDLARLDAHRPLELCRVDLLALASDAADHARAMDPKRTITLEVLDGPGTPEVLGDESRLRQVLRLVANAIQHTPESADVTJRVGTEGDAOEADDGPGMSQEDALRVFERFYRADSSRARASGGTGCLSLIVDSLVAHGGAVTAVTALGEGCCFRVSLPRVSDVDQLSRTLTPVVPGPP
Appendix B

*phoP* and PhoP Sequences

**Gene sequence**

*M. tuberculosis* H37Rv|Rv0757

*phoP*: 744 bp

```
1  atgcggaaag ggttgatct cgtgacggcg ggaaccccag ggcggaaacac cacacgggag
61  gctcgtgtcc tcgtggtcga tgatgaggcc aacatcgttg aactgctgtc ggtgagcctc
121  aagttccagg gtttcgcag ctacaccccg accaaggggg cacagccgct ggtcggcggcccc
181  cgggaaaccc ggccggacgc ggtgatcctc gatgtgatga tgcccgggat ggacggcttt
241  ggggtgctgc ggcggctgcg cgccgacggc atcgatgccc cggcgttgtt cctgacggcc
301  cgtgactcgc tacaggacaa gatcgcgggt ctgaccctgg gtggtgacga ctatgtgaca
361  aagcccttca gtttggagga ggtcgtggcc aggctgcggg tcatcctgcg acgcgcgggc
421  aagggcaaca aggaacacag taatgttca ctgaagtttg ccgatatcga gctcgacgag
481  gagacccacag aagttggcag ggcgggcca gatcgcgggt ttcgcccac caaatgacctcc
541  tcgtcggact atttctgatt caacccggtcg agctgcgctg gcaacgtccaa gattctgac
601  cacgtgggct gaagctgccg ccgggttcgc ctgcgctcct ccatcgatt ccgtcttcgc
661  ctgcacgctg agatgacacg ggcgggaggg gtgcttgtgc acaagacgcc acaactgcg ccggggtgggcc
721  tacgtactgc gggagcctcg atga
```

**Protein sequence**

*M. tuberculosis* H37Rv|Rv0757

PhoP: 247 aa

```
MRKGVDLVTAGTPGENTTPPEARVLVVDEANIVELLSVKFQGFIVYATNGA
QALDRARETRPDAILVMMPMDGFGLRLRRLADGIDAPALFLTARSDLQDKI
AGLTLGGDDYVTKPSLEEVVARLRLRRAGKGNKPRNVRRLTFADIELDEETH
EVWKAGQPVSLSPEFTLLRYFVINAGTVLSPKILDHVWRYDFGGGDNVVESY
VSYLRRKIDTGEKRLLLHTLRGGYVLRPR
```
Appendix C

*M. smegmatis* and *M. tb mshC* and MshC sequences

Gene sequences

*M. smegmatis* MC2-155
cysS: 1239 bp

1  atgcaatcgt gtcggcacc gccttcgcc gacagtgccct cgctgtgcgc
61  ctcttcgaca gcgctgatcg ccaggtccgg cccgtcacac cgggaccgac cgcaaccatg
121  taactgtcgc gcgcgaacgc accattggc gacggtgctg atgcagttctg
gcaggtgcac gcgcgaacgc accattggc gacggtgctg atgcagttctg
181  tcacgcggtc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc
241  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
301  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
361  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
421  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
481  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
541  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
601  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
661  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
721  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
781  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
841  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
901  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
961  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
1021  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
1081  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
1141  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
1201  cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc cggcgccggc aacgtgcacc
**M. tuberculosis** H37Rv|Rv2130c

**mshC:** 1245 bp

```
1  atgcagctg  ggtattgccc  accggttccg  gtgttgccgg  gacgaggccc  gcagctacgg
61  ctgtacgaca  gcgccgaccg  gcaggtccgt  ccggtggcgc  ccggatctaa  ggccaccatg
121  tacgtctgcg  ggatcacgcc  ctacgacgcc  acgcatctgg  gccatgctgc  cacctatgtg
181  acgttc
241  atgcagctg  ggtattgccc  accggttccg  gtgttgccgg  gacgaggccc  gcagctacgg
301  cggatctctg  atgcagctg  ggtattgccc  accggttccg  gtgttgccgg  gacgaggccc  gcagctacgg
361  ctacgacgcc  acgcatctgg  gccatgctgc  cacctatgtg
421  ggtattgccc  accggttccg  gtgttgccgg  gacgaggccc  gcagctacgg
481  tacgtctgcg  ggatcacgcc  ctacgacgcc  acgcatctgg  gccatgctgc  cacctatgtg
541  ctacgacgcc  acgcatctgg  gccatgctgc  cacctatgtg
601  ggtattgccc  accggttccg  gtgttgccgg  gacgaggccc  gcagctacgg
661  ctacgacgcc  acgcatctgg  gccatgctgc  cacctatgtg
721  cggatctctg  atgcagctg  ggtattgccc  accggttccg  gtgttgccgg  gacgaggccc  gcagctacgg
781  ctacgacgcc  acgcatctgg  gccatgctgc  cacctatgtg
841  ctacgacgcc  acgcatctgg  gccatgctgc  cacctatgtg
901  atgcagctg  ggtattgccc  accggttccg  gtgttgccgg  gacgaggccc  gcagctacgg
961  ctacgacgcc  acgcatctgg  gccatgctgc  cacctatgtg
1021  ggtattgccc  accggttccg  gtgttgccgg  gacgaggccc  gcagctacgg
1081  ctacgacgcc  acgcatctgg  gccatgctgc  cacctatgtg
1141  ggtattgccc  accggttccg  gtgttgccgg  gacgaggccc  gcagctacgg
1201  ggtattgccc  accggttccg  gtgttgccgg  gacgaggccc  gcagctacgg
```

**Protein sequences**

*M. smegmatis* MC2-155

MshC: 412 aa

MQSWSAPAIPVYGPGRPLALRLRLDADQVRPVTGPTATMYVCGITIPYDAIHTLGHAAATYTLFDLVHRLWLDAGHTQYVQNYTVDPLFRAERDILDWRTLGDR
ETQLFREDMAALRVLPPHDYVAATDAEVEVEMVEKLLASGAAYVEDAEYPD
VYFRADATAFQGYESEGRYDRDDMTLFAERGDPDRPGKSDLALLWRAERP
EPSWPSPFGRGRPGWHECSAIALTRITGLDIGQGGSIDLFPHHEYSAIAHESVT
GERRFARHYVHTGMIWGDHMSKSRGNLVLSQLRAQGVDPSAIRGLFGS
HYREDRFWSNENLDEANRLARWRSATAPADTVIARVRQYLADDLDTDP
KALAAALDGWCTDALSYGGHDTESPRVATTVDALLGVDL
M. tuberculosis H37Rv\textbackslash Rv2130c
MshC: 414 aa

MQSWYCPPPVLPGRGPQLRLYDSADRQVRPVAPGSKATMYVCGITPYDATHL
GHAATYVTFLHRLWLDLGHHELHYVQNI\textbackslash DDPLFERADFGRVDWDLQA
EVALFEDMAALRVLPPQDYVGATEAIAEMVELIEKMLACGAAYVIDREMGEY
QDIYFRADATLQFGYESGYDRDTMLRLCEERGGDPRRPGKSDELDAWLRAAR
PGEP\textbackslash WPSPSFGPGRPGWHVECAAIALSRI\textbackslash SGSLIDIQG\textbackslash SLDLIFPH\textbackslash EFTAAHAEC
VSGERRFARHYVHAGMIGWDGHKMSRG\textbackslash N\textbackslash LVLV\textbackslash SALRAQDVEPSAVR\textbackslash LGLL
AGHYRADRF\textbackslash WSQQVLDEATARLHR\textbackslash WTAT\textbackslash LPA\textbackslash GPA\textbackslash AVDVVARVRR\textbackslash YLADDL
DTPKAIAALDG\textbackslash WVTDAVE\textbackslash YGGHD\textbackslash AGPKLVAAT\textbackslash IDALLGVDL