TRIGGERED FOLDING OF A *MYCOBACTERIUM TUBERCULOSIS* ADHESIN, HEPARIN BINDING HEMAGGLUTININ ADHESIN (HBHA).

Joseph Vincent Lomino

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry and Biophysics.

Chapel Hill

2011

Approved by:

Advisor: Professor Matthew Redinbo, Phd

Reader: Professor Edward Collins, Phd

Reader: Professor Jian Liu, Phd

Reader: Professor Ann Matthysse, Phd

Reader: Professor Gary Pielak, Phd

Reader: Professor Richard Wolfenden, Phd

ABSTRACT

Joseph Vincent Lomino: Triggered folding of a *Mycobacterium tuberculosis* adhesin, heparin binding hemagglutinin adhesin (HBHA).

(Under the direction of Matthew R. Redinbo)

The heparin-binding hemagglutinin (HBHA) is a surface adhesin on human pathogen *Mycobacterium tuberculosis*. Previously, it has been shown that HBHA exists as a dimer in solution. We investigated the detailed nature of this dimer using circular dichroism spectroscopy and analytical ultracentrifugation techniques. We demonstrate that the heparan sulfate (HS) binding region does not play a role in dimerization in solution, while the linker region between the predicted N-terminal coiled-coil and the C-terminal HS binding region does impact dimer stability. The majority of contacts responsible for dimerization, folding, and stability lie within the predicted coiled-coil region of HBHA, while the N-terminal helix preceding the coiled-coil appears to trigger the folding and dimerization of HBHA. Constructs lacking this initial helix or containing site-specific mutations produce non-helical monomers in solution. Thus, we show that HBHA dimerization and folding are linked, and that the N-terminal region of this cell surface adhesin triggers the formation of an HBHA coiled-coil dimer.

DEDICATION

I would like to dedicate this work to those who could not be here to see this moment, but have contributed to my education and my development as a person. In loving memory:

Joseph “Duke” Tona

Joseph Lomino, Jr.

Joseph Lomino, Sr.

Harold H. Siegel

ACKNOWLEDGEMENTS

I would like to begin with acknowledging my wife, Lee Lomino. Who in truth is the ultimate driving force in my graduate school career. It is only due to her love, understanding, and patience that I stayed the course and continued with my education. There are not enough pages in my thesis to express my appreciation and love for her.

Secondly, I need to thank my mother, Rose C. Charpentier. Since the very beginning, my mom has taken a proactive role in my education. Her interest, coupled with her constant love and encouragement made for an environment where I wanted to learn and succeed. There were many times when I was younger that my mom fought to have me put in the accelerated classes. One time she drove me the superintendants office, when we were living in New York City, to take an entrance exam for the science and math program. I was eleven years old sitting in his office taking a two hour test. It is one of the moments that I hope to replicate in spirit for my own children.

I would like to acknowledge my grandmothers, Grace Lomino and Antionette Tona. I would not be nearly as motivated or educated if not for years of encouragement and patience. I still remember reading books to my grandmother Antionette on my bed as a child and working on phonics flashcards and multiplication tables with my grandmother Grace.

I would be remiss not to thank my committee members, current and previous, for all there help through graduate school. First, Dr. Braunstein for supplying me with materials and helping me to better understand *Mycobacterium tuberculosis*. Dr. Collins for his constructive criticisms

of my work and presenting ability that ultimately made me a better scientist and presenter. Dr. Wolfenden, whose early mentorship helped rekindle my love of science. Dr. Ann Matthysse, who first introduced me to microbial adhesion and helped me start my first project in the Redinbo laboratory. I had many interesting conversations with Dr. Pielak about my project or just about science in general. I looked forward to every scheduled meeting over coffee. I have learned a lot from Dr. Pielak and thank him for all his time. Lastly, I need to thank Dr. Liu for his interest in my work and my career. Dr. Liu's interest and zeal for science, in particularly Glycobiology, cultivated my interest in his field, which ultimately is the direction I am choosing to pursue. In hind-sight, it is funny how typing "UNC Heparin" into a search engine led me to Dr. Liu and how he then helped my get a postdoctoral position with Dr. Lai – Xi Wang.

An often unheralded member of many theses is Ashutosh Tripathy. Without Ash there is no HBHA paper. His instruction on many biophysical methods was the catalyst for my experiments, which led to the watershed moment in my thesis. I am in his debt.

Throughout my time in the Redinbo laboratory there have been many inmates at the asylum. My first partner in crime was Scott Lujan. Scott and I came up with "creative" ways of approaching scientific problems, which ultimately turned into Rube Golderg type schemes. Scott's way of thinking about problems stayed with me long after he left the lab. My second partner in crime was Yuan Cheng. Throughout the course of the HBHA project I could tell I was making headway if I was able to convince Yuan about my data. His encouragement and criticisms were much needed and appreciated over the years. I would also like to acknowledge, Monica Frazier for being a great friend, lab mate, and proof reader of important emails. My time in the Redinbo laboratory was more enjoyable with her around. I also would like to thank

Michael Johnson, Li Pan Yao, Keith Ballentine, Lisa Withers, Michael Miley, and Rebekah Nash.

Finally, I would like to thank my P.I., Dr. Matthew Redinbo. I came to UNC – Chapel Hill to work with Dr. Redinbo after working with his associates. It was through his mentoring that I was able to become an independent researcher. I thank him for all the time I spent in his laboratory.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A.A.	Amino Acids
ABRC	Arabidopsis Biological Resource Center
Abs	Absorbance
AFM	Atomic Force Microscopy
ATP	Adenosine-5'-triphosphate
AUC-SE	analytical ultracentrifugation sedimentation equilibrium
BCG	Bacillus Calmette - Guerin
CC	Coiled-coil
CCD-NKI	Crystallization Construct Designer - Netherlands Cancer Institute
CD	circular dichroism
CFP-10	culture filtrate protein
CPD	cysteine protease domain
DevS	differentially expressed in the virulent strain
DTT	Dithiothrietol
EK	Enterokinase
ESAT-6	early secreted antigenic target
GAF	cGMP-specific phosphodiesterases, adenylyl cyclases and FhlA
GDP	guanosine diphosphate
HBHA	Heparin Binding Hemagglutinin Adhesin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human Immunodeficiency Virus
HK	Histidine Kinase

HS	heparan sulfate
HSP	Heat Shock Protein
IFN - γ	intereferon gamma
IL-12	interleukin 12
iNOS	inducible Nitric Oxide Synthase
Ins6p	inositol-6-phosphate
IPTG	isopropyl-β-D-thiogalactopyranoside
kDa	kilo Dalton
KF	Potassium Fluoride
LIC	Ligation Independent Cloning
LM	lipomannan
M1P	mannose – 1 – phosphate
M6P	mannose – 6 – phosphate
M6PR	mannose – 6 – phosphate redutase
ManLam	lipoarabinomannan
MARTX	Multifunctional Autoprocessing Repeats-in-Toxin
MBP	Maltose Binding Protein
MDR	Multi-drug resistance
MR	Mannose Receptor
MRE	mean residue ellipticity
Mt. Manb	Phosphomannomutase
Mtb	*Mycobacterium tuberculosis*
MWCO	Molecular Weight Cut Off

NaCl	Sodium Chloride
NADP	nicotinamide adenine dinucleotide phosphate
NLR	nucleotide-binding oligomerization domain like receptors
nm	nano meter
NO	Nitric Oxide
NOD-	nucleotide-binding oligomerization domain
NRMSD	normalized root mean square deviation
NRP	non-replicating persistence
NTA	nitriloacetic acid
O.D.	Optical Density
^{o}C	degree Celsius
PAS	period circadian, aryl hydrocarbon receptor nuclear translocator protein, single-minded protein
PDB	Protein Data Bank
r	radius
RD-1	regions of difference 1
RIPL	Arginine, Isoleucine, Proline, Leucine
RPM	Revolutions per minute
RR	Response Regulator
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
Sp-A	surfactant protein A

SPR	surface plasmon resonance
TB	Tuberculosis
TCEP	tris(2-carboxyethyl)phosphine)
TEV	Tobacco Etch Virus
TLR	Toll-like receptors
T_m	Melting Temperature
UV	Ultraviolet
XDR	extreme drug resistance

Chapter 1:

A Record of Tuberculosis

Tuberculosis has been a human plague for centuries. People infected with Tuberculosis would suffer from fevers and severe weight loss contributing to the appearance of being consumed from the inside, hence the disease original name, Consumption. On March 24th 1882, Robert Koch determined that *Mycobacterium tuberculosis* was the etiological agent of Tuberculosis[1]. In his work he demonstrated how the disease was transmitted and suggested patients be quarantined to prevent further spread of the disease. Dr. Koch also developed methods for staining and growing the mycobacterium, to assist in the study of the disease.

It is estimated that one third of the world's population are infected with *Mycobacterium tuberculosis* and there are 9.7 million new active cases of TB (new and relapse) per year[2]. There is resurgence in TB cases in both the developed and developing countries, amongst immunocompromised populations, such as those infected with HIV, the elderly and the newly born. Treatment of active Tuberculosis consists of a 6 to 9 month regimen of four anti-microbial drugs; rifampin, isoniazid, pyrazinamide and either ethambutol or streptamycin. These combination, of drugs are used to prevent the occurrence of antibiotic resistance. Multi-drug resistance (MDR) and extreme drug resistance (XDR) cases of Tuberculosis are becoming more prevalent especially in the developing countries where antibiotics are available without prescription and patients are less likely to finish the treatment[3]. Finding new classes antibiotics,

to counteract the growing number of MDR strains of *Mycobacterium tuberculosis* is highly important.

Disease Progression

Tuberculosis progresses through several stages of infection. The droplet nuclei, containing the bacterium, are inhaled by the human host allowing the bacilli access to the lungs and where infection is initiated. There is still some debate about the role lung epithelial cells play in infection or dissemination of the bacteria, but greater than 90% of the alveolar surface are pneumocyte cells, increasing the likelihood of mycobacterial – epithelial cell interactions[3]. It is here that the alveolar macrophages ingest the bacterium, but since the macrophages are not yet activated they are unable to destroy the mycobacterium. A localized pro-inflammatory response is triggered, recruiting mononuclear cells to the site of infection. The bacteria that escape from the non-activated macrophages are phagocytosed by activated macrophages or dendritic cells. The mycobacteria interact with many possible receptors on these immune cell surfaces, including Toll-like receptors (TLR), nucleotide-binding oligomerization domain (NOD-) like receptors (NLRs), C-type lectins (including the mannose receptor [MR]), and other complement, scavenger, surfactant protein A (Sp-A), and cholesterol receptors[3]. The receptor through which M. tuberculosis is internalized dictates bacterial treatment. Interaction with the mannose receptor (MR) on the macrophage surface, with lipoarabinomannan (ManLam) structures on the bacteria has been shown to prevent maturation of the phagosome with the lysosome[4]. The bacterium is able to survive in a vacuole within the macrophage. This topic will be further discussed in chapter 3. Phagocytosis via the TLRs is the main extension of the innate immune response triggering a pro-inflammatory response. The cellular immune response is linked in

particular to the production of IFN - γ. In most cases, the infection is contained by walling off the infected site, forming a granuloma[5]. In some cases the resistant bacilli can survive inside macrophage, and go into a dormant state. At this point the host is categorized as having latent Tb. The triggering of the dormant state has been studied and one of the proposed mechanisms is discussed in chapter 4. The dormant bacteria can exist in the granuloma for the life of the host without the host presenting any symptoms of the disease.

In those cases where the disease progresses or "awakens", a region of genes named, regions of difference 1 (RD-1) are up-regulated[6]. RD-1 genes are present in all virulent strains of *M.tb*, but absent from non-pathogenic strains[6]. Two proteins which assist in mycobacterial escape are the early secreted antigenic target (ESAT-6) and culture filtrate protein (CFP-10). ESAT-6 has been shown to insert into membranes causing cytolysis of macrophages, demonstrating a possible mechanism for escape from the macrophage and release into the extra-cellular milieu[7]. Under these circumstances the granuloma center undergoes caseation releasing the bacteria into the lung environment. The infection progresses to other regions of the lung or escapes the lung environment. One method of escape is through host sputum spread by productive coughing. The bacteria can also gain access to the blood stream and disseminate to other organ systems. While the mechanism of dissemination still needs to be elucidated, the effect of particular proteins on the process has been studied. A protein important for dissemination from the lungs, heparin binding hemagglutinin adhesin (HBHA), is discussed in chapter 2[8]. Due to its significant impact on public health, the pathogenesis of *Mycobacterium tuberculosis* has been studied for many years. The focus of this work was on Heparin Binging Hemagglutinin Adhesin, Phosphomannomutase (Mt. Manb), and DevS. These proteins facilitate adhesion, dissemination, evasion, and hibernation.

Chapter 2:

Triggered *Mycobacterium tuberculosis* Heparin Binding Hemagglutinin Adhesin Folding and Dimerization

Joseph V. Lomino[1, 2], Ashutosh Tripathy[2, 3], and Matthew R. Redinbo[1, 2, 4, 5, 6, *]

Departments of Chemistry[1], Biochemistry and Biophysics[2], UNC Macromolecular Interactions Facility[3] and Department of Microbiology and Immunology[4], Program in Molecular Biology and Biotechnology[5], and the Lineberger Comprehensive Cancer Center[6], University of North Carolina, Chapel Hill 27599-3290

Published in the Journal of Bacteriology (Pending)

Introduction

When *Mycobacterium tuberculosis* is inhaled as an aerosol, infection starts in the lung environment. Upon entering the lungs there is a higher likelihood of the mycobacterium encountering the lung epithelial cells than an alveolar macrophage. The mycobacteria need a complement of adhesins to interact with this cellular environment along with macrophage interactions. One of the adhesins that *Mycobacterium tuberculosis* uses to interact with these epithelial cells is heparin binding hemagglutinin adhesin (HBHA)[9]. Genes encoding for HBHA are only found in pathogenic strains of mycobacteria, *Mycobacterium leprae, Mycobacterium tuberculosis, and Mycobacterium avium,* not in non-pathogenic strains such as *Mycobacterium smegmatis*[10]. HBHA is known to interact with the polysaccharide Heparan Sulfate (HS), which extends from some epithelial cell receptors[9]. The role of HBHA in pathogenesis is still debated, but its function in dissemination from the lungs has been shown[8]. HBHA has been characterized with an ability to agglutiniate red blood cells while also auto-aggregating mycobacteria[9]. HBHA has also been seen to interact with actin molecules within the epithelial cells causing rearrangement of the cytoskeleton, which is hypothesized as the mechanism for escape of the mycobacteria from the lung environment[11]. HBHA has also been investigated as a potential diagnostic tool of latent tuberculosis[12], as a possible vaccine[13], and boosts BCG immunity, probably by inducing cytokines that may help with production of Th1 effector-memory lymphocytes[14].

HBHA is a 199 amino acid (a.a.) long protein found on the surface of Mycobacterium tuberculosis. The adhesin is not evenly distributed among the surface, but is localized to the side facing the epithelial cells[15]. The mechanism by which HBHA is anchored to the surface of the

cell is unknown, but there is some thought that the glycosylation on the N-terminal asparagine might act as an anchor for the protein. The saccharide composition is 41.6 % glucose, 17% xylose, 10.2 % mannose, arabinose 8.8%, and 9.2% galactose contributing to 2.8 % of the total mass of HBHA [9].

There has been much work done on characterizing the oligomerization state of HBHA in vivo and in vitro. Dupres *et al*. unfolded surface associated HBHA using Atomic Force Microscopy (AFM), and concluded that HBHA was a monomer based upon the force of unfolding along with the length of the unfolded molecule[15]. Other work has shown HBHA to be an obligate dimer in vitro[9,16]. It is also predicted from the agglutination data that HBHA exists in a higher ordered oligomerization state[9]. From primary sequence analysis, the presence of a coiled-coil region present from residues 25 to 108 was detected. Coiled-coil (CC) motifs are known regions of protein – protein interactions. These regions are characterized by a heptad repeat (a-b-c-d-e-f-g), where residues a and d are hydrophobic residues, e and g, are charged residue and b, c and f are hydrophilic residues[17]. Coiled-coils are alpha helices that coil around each other, two to seven strands, where the hydrophobic residues are buried in a core and the hydrophilic residues are exposed.

Work done by Delogu and Brennan identified the C-terminal region as the Heparan Sulfate binding region of HBHA using heparin as a model ligand[18]. Heparin is a highly sulfate form of HS used as a model in binding studies since it mimics the highly sulfate regions of HS. Residues 161 to 199 consist of lysine, alanine, and proline repeats, which present a localized positive charge that are complementary to the negatively charge sulfate groups of HS. Removal of this basic region, causes a loss of heparin binding ability of HBHA, while fusion of the C-terminal region to Maltose Binding Protein (MBP), supports MBP binding to heparin[19].

The primary focus of this work was to understand the contribution of particular regions of HBHA on folding, dimerization, and stability. Methods used in the investigation of HBHA were circular dichroism (CD) wavelength and temperature studies and analytical ultracentrifugation sedimentation equilibrium (AUC-SE).

Construct Design:

With no atomistic information concerning the secondary structure of HBHA, software was used to give a starting point. HBHA was put through secondary structure prediction server (CCD-NKI) and the resulting data was aligned with previously published coiled-coil predictions of HBHA (Figure1) [20]. From analysis of the alignment, we chose regions to investigate based upon known function and secondary structure prediction. Residues 161 to 199 were designated the HS binding region, while residues 25 to 110 is the coiled-coil region. The coiled-coil region is predicted to end at residue 108, but the alpha helix is also predicted to continue until residue 110, which is then followed by a predicted gap in structure. Amino acids between the coiled – coil and the HS binding region was designated the linker region, while any residue N-terminal to the coiled – coil is part of the N-terminal region. By breaking the protein down into parts we were able to investigate how each section functions. The list of constructs, used in the in the experiments are in Table 1.

Most constructs started at residue 6 (aspargine) versus native start at residue 1 (methionine). The rationale for the alternate start concerned the half-life of the constructs in solution. Constructs that started at residue 6 were more stable, which was important when performing analytical ultracentrifugation sedimentation equilibrium techniques that were run for three days. In choosing the construct end residues, we looked at both the end of predicted

secondary structure along with organization of the HBHA. To test each regions contribution to secondary structure, stability, and folding, residues 199 (lysine), 156 (glycine), 111 (serine), 99 (valine) and 88 (glutamate) were chosen as the C-terminal ends of the constructs. Along with residue 6, residue 20 was used as the N-terminal start so the contributions of the N-terminal region could be ascertained. Residue 20 was chosen over residue 24 as the start, because residue 24 was part of predicted secondary structure. The resulting data from a construct starting at 24 could be due to the disruption of the predicted alpha helix versus loss of the N-terminal region. Upon further analysis of the secondary structure and coiled-coil alignment a hydrophobic stretch of residues 12 to 18 lies within a predicted alpha helix. In previous work with coiled-coils, it was shown that there are motifs that control the formation of coiled-coils, called trigger sequences. As part of our study, we investigated the possibility of the N-terminal helix as a trigger sequence.

Purification:

The constructs created with the aforementioned starts and stops along with mutations were amplified from *M. tuberculosis* (strain H37Rv) genome via polymerase chain reaction and incorporated into the LIC (ligation independent cloning) vector pMCSG7[21]. This vector incorporates a hexa-his tag at the N-terminus of the HBHA construct linked via tobacco etch virus (TEV) protease cleavage site. L14A, L15A mutants of HBHA 6_199 and 6_111 were introduced using site-directed mutagenesis. Sequence confirmed constructs were inserted into pMCSG7 then transformed into chemically competent *E. coli* BL21 (DE3) AI cells (Invitrogen). Transformed cells grew to an optical density of 0.8 in terrific broth (TB) at 37^{o}C, with shaking. An L-Arabinose solution was added to the growth at a final concentration of 0.2% v/v, and the temperature was reduced to 18^{o}C for 30 minutes. Protein expression was induced with 0.5 mM

IPTG (isopropyl-β-D-thiogalactopyranoside) and cells were allowed to grow for 16 hrs. Cells were then centrifuged at 4500 *xg* for 30 minutes at 4°C. Pellets were suspended in 50 mL lysis buffer (50 mM potassium phosphate, 500 mM NaCl, 5% glycerol, and 20 mM imidazole, pH 7.4) per 4 liters of culture. Suspended pellet was frozen drop-wise into a liquid nitrogen bath. Pellets were resuspended in the lysis buffer plus protease inhibitor tablets (Roche) and DNase. Lysis was achieved using a Branson Sonic Dismembrator at 30% power for 3 cycles of 1 minute at 4°C. Constructs were purified over a HisTrap column (GE Healthcare). The hexa-his affinity tag was removed using TEV protease with further purification over a Superdex 200 column (GE Healthcare) pre-equilibrated in CD buffer (10 mM Potassium Phosphate pH 7.5, 200 mM KF). Fractions containing HBHA were combined and used in the following experiments.

Circular Dichroism:

Circular dichroism is a measure of the difference in absorbance of chiral (optically active) molecules of plane polarized light consisting of two components, left (counter-clockwise) and right (clockwise). The measurement of the difference: $\Delta Abs = Abs_L - Abs_R$ with ellipticity (θ) being the unit of circular dichroism results. The relationship of ellipticity (θ) to absorbance, $\theta = 32.98\,\Delta A$, with the final CD spectrum plotted as ellipticity as a function of wavelength. We investigate the secondary structure of proteins by viewing the absorbance spectra from approximately 240 to 180 nm. Within this range the absorption is principally from the peptide bond, but each secondary structure element generates distinct spectral bands. Helices produce broad negative bands with a center around 222 nm and a negative and positive couplet signal around 208 nm and 190 nm, respectively[22]. The beta-sheet signal is characterized by a negative band at 215 nm and a positive band at 195 nm. If the protein in question contains a

combination of secondary structure then the resultant signal will be an ensemble of the signals. Circular dichroism is also sensitive to changes in secondary structure of proteins. When there is a loss of secondary structure, a blue shift toward 200 nm is seen in the CD spectrum due to an increase in random-coil signal. Spectral data of target proteins can be analyzed using a variety of algorithms and data sets. Common algorithms are SELCON, CDSSTR, and CONTIN/LL, which can be accessed using the online program, Dichroweb [23]. Results from data analysis give estimates on percentages of the types of secondary structure present in the target protein.

HBHA constructs were analyzed by circular dichroism spectroscopy. Wavelength spectra were collected in CD Buffer (10 mM potassium phosphate pH 7.5 and 200 mM potassium fluoride) from 260 to 185 nm using an Aviv 62 DS CD Spectrophotometer. HBHA constructs were studied at a concentration of 10 µM in quartz cuvettes (Hellma) with a path length of 0.1 cm. Each sample was scanned three times at 20^{o}C with a 10 second averaging time. Spectra were averaged and buffer signal subtracted. Values were converted from ellipticity to mean residue ellipticity (MRE) (deg * cm^{2} * dmol^{-1}) using equation:

$$[\theta] = (\theta * Mr * 0.1)/(C * l) \quad (1)$$

where θ is ellipticity in mdegs, *Mr* is the molecular weight divided the number of peptide bonds, *C* is the concentration in g/L and *l* is the path length in cm.

HBHA construct stability was analyzed via thermal denaturation by heating the samples from 4^{o}C to 85^{o}C using 1 degree steps with an equilibration time of 2 minutes for each temperature step. Measurements were taken at 222 nm with a 10 second averaging time for each degree step within a 0.3 degree dead band. Transition curves were normalized to fraction folded using the equation:

$$F = ([\theta]_{obs} - [\theta]_u)/([\theta]_n - [\theta]_u) \quad (2)$$

where $[\theta]_{obs}$ is the observed signal at a given temperature, $[\theta]_u$ is the value when the peptide is unfolded, and $[\theta]_n$ is the value when the peptide is in its native state. Reversibility was checked by rapid cooling to 25°C, followed by a wavelength scan collected from 260 to 185 nm.

Secondary structure of each construct was further analyzed using three different algorithms; CONTIN/LL, CDSSTR, and SELCON. Two reference sets, set 4 and set 7, (appendix A) contained the broadest sampling of proteins at wavelengths pertinent to our studies. Wavelength data for each construct, between 240 nm to 190 nm, were uploaded to the Dichroweb website[23]. Each construct data was analyzed by different combinations of algorithm with each selected reference set. Fit integrity was evaluated using normalized root mean square deviation (NRMSD) and plotting experimental data versus calculated data. The product of the analysis is percent helix, beta sheet, turn, and unstructured.

Circular Dichroism Spectroscopy: Wavelength Spectra

CD spectra from 260 nm to 185 nm for each construct at 10 μM are shown in Figures 2a-d. The spectra of HBHA 1_199 and 6_199 were identical, indicating that the truncation of the first five residues did not affect the overall structure of the protein (Figure 2a). The CD spectra of constructs 6_156 and 6_111 were then compared to that of 6_199, which showed that the secondary structure of HBHA became dominated by alpha helices as the predicted coiled-coil region was isolated. The spectrum from HBHA 6_138 was similar to HBHA 6_156, demonstrating that cutting into the linker region had little effect on overall secondary structure composition (Figure 2b). The CD spectra from HBHA 6_88 and HBHA 6_99 were similar to one another, and both showed loss of alpha helical secondary structure due to truncation from the

C-terminus. Compared to HBHA 6_111 (Figure 2b), the minima at 208 nm and 222 nm are more positive and the maximum at 190 nm is more negative. These transitions were accompanied by a blue shift of the minimum observed near 208 nm; together, these data support the conclusion alpha helical secondary structure is lost in these constructs relative to longer version of HBHA examined.

Previous published work had shown that truncation of the first 24 residues of HBHA results in a loss of secondary structure[16]. Therefore, a construct starting at residue 20 (alanine) was designed, since a truncation at residue 25 is within a predicted helix and close to the predicted coiled-coil start. The introduction of the N-terminal truncation caused a loss of secondary structure when compared to 6_199, as indicated by the observed shift in the minimum at 208 nm to 200 nm and a loss of CD signal intensity at 222 nm and 190 nm (Figure 2c). At the N-terminus of HBHA, there is a marked stretch of hydrophobic residues (see Fig. 1). Two leucines within this region, leucines 14 and 15, were mutated to alanines using site-directed mutagenesis with the goal of disrupting the coiled-coil. The mutation of leucines 14 and 15 to alanines had a similar disruptive effect on the secondary structure (Figure 2c). Thus, altering the N-terminal region lying outside the predicted coiled-coil has adverse effects on the alpha helical secondary structure of HBHA

The leucine mutations were then incorporated into the most alpha helical construct, HBHA 6_111. However, following TEV protease cleavage during purification of the construct, the protein formed a soluble aggregate as determined by size exclusion chromatography and dynamic light scattering (data not shown). This aggregate was examined on the CD spectrophotometer along with HBHA 6_111 and HBHA 20_199 (Figure 2d.). We observed a total loss of secondary structure signal when the leucine mutations were introduced in the 6_111

construct relative to 6_199 due to aggregate state of construct. Taken together, these data demonstrate that the primary structural component for HBHA is the alpha helix and the maintenance of this structure is insensitive to C-terminal truncations up to the predicted end of the coiled-coil at residue 111, but highly sensitive to mutation or truncations within the N-terminus leading up to residue 20.

Dichroweb Output

Decovolution software has been shown to work best for helical proteins[23]. Each constructs secondary structure content were evaluated using CONTIN/LL, SELCCON, and CDSSTR along with reference sets ref 4 and ref 7 (Appendix A). Using various software and reference sets secondary structure values were generated for each construct. Fitting results are listed in Table 2. Values are estimates of secondary structure present in the sample, not a precise quantification of the structure. The most alpha helical signal was obtain from HBHA 6_111, with HBHA 20_199 and HBHA 6_199 L14A L15A having the least alpha helical content. As HBHA is truncated from the C-terminus, there is an increase in alpha helical structure until we cut into the predicted coiled - coil (HBHA 6_88).

Thermal Denaturation

The CD signal at 222 nm is generally accepted as an accurate measure of the alpha helical content of a protein and because HBHA is composed of primarily alpha helices, 222 nm was considered an appropriate measure of HBHA folding and stability[24]. The constructs 6_199, 6_156, and 6_111 were analyzed along with the full length construct, HBHA 1_199 (Figure 3a). Melting temperatures (T_m) were determined and are listed in Table 3. As reported above, the signal generated by HBHA 1_199 was similar to that of 6_199 (Figure 3a). No difference in

protein stability was detected upon the removal of the HS binding domain (*e.g.*, see the data for HBHA 6_156 in Figure 3a). However, the T_m began to decrease once the full linker region was removed from the HBHA construct (*e.g.*, HBHA 6_111 in Fig. 3a). Truncating the linker region increased the alpha helical secondary structure and had an effect on stability as seen by a decrease in T_m (Fig. 3b). Once the predicted coiled-coil region was disrupted, significant reductions in T_m were observed (Table 3). It was noted that when transitioning from 60^{o}C to 40^{o}C the data for HBHA 6_88 were noisier relative to the other constructs (Figure 3b), likely due to a more disordered nature of the construct.

The change in T_m as a function of peptide length is plotted in figure 3c. As residues are truncated from the C-terminus, only a small decrease in the T_m of the HBHA constructs is observed. However, as we disrupted the predicted coiled-coil region, a more significant decrease in T_m is observed (see also Table 3). The limited stability of constructs HBHA 20_199, HBHA 6_199 L14A L15A, and HBHA 6_111 L14A L15A did not allow for successful analysis of these proteins by thermal denaturation. Collectively, though, these data indicate that disrupting the coiled-coil domain from the C-terminus weakens the fold of the HBHA dimer.

Analytical Ultracentrifugation Sedimentation Equilibrium:

Sedimentation equilibrium experiments were performed on HBHA constructs at 20° C in a Beckman XL-I analytical ultracentrifuge using 6-sector cells and an An50Ti rotor. Samples at three different concentrations (20, 10, 5 μM) were spun at three velocities (3.22 x10^4, 7.24 x10^4, and 1.30 x10^5 *xg*) until equilibrium was reached. Absorbance was monitored every two hours at 230nm and 280 nm. Absorbance offset values were determined using the meniscus depletion method by overspeeding the samples at 1.63 x10^5 *xg* for 6 hours. The program SEDNTERP

(John Philo, Thousand Oaks, Ca and RASMB) was used to calculate solvent density and viscosity along with molecular mass and partial specific volume for each HBHA construct. Data sets were initially analyzed using SEDFIT [25] in order to generate the proper file type for analysis with SEDPHAT[26]. Using SEDPHAT, a species analysis model was applied to each construct data set to determine the solution make up for each construct. Standard curves for each HBHA construct were generated at 230 nm in order to derive the extinction coefficients using the Beer-Lambert law. Data for each construct was also plotted using the equation:

$$\ln\left[\frac{C_r}{C_{r_0}}\right] = \frac{M(1-\bar{\nu}\rho)\omega^2}{2RT}\left(r^2 - r_0^2\right) \quad (3)$$

where C_r is the concentration at radius, r, C_{r0} is the concentration at the reference point, M is the molecular weight, $\bar{\nu}$ is the partial specific volume of the molecule, ρ is the density of the solution, ω is the radial velocity of the experiment, R is the gas constant, T is the temperature of the experiment (K), r is radius and r_0 is the reference radius. The resulting plot is linear if the molecule population is monodisperse, while a curved lined indicates the polydispersity of the sample. The slope of the line is proportional to the mass of a single species.

Because HBHA may form an elongated dimer, we employed analytical ultracentrifugation sedimentation equilibrium (AUC-SE) to obtain accurate masses of HBHA constructs due to the shape independence of this technique[27]. The full length construct of HBHA 1_199 was compared to HBHA 6_199 to confirm that the deletion of the first 5 residues had no effect on oligomerization (Figures 4a – f). Regional constructs HBHA 6_156 and HBHA 6_111 were examined at 3 different concentrations, with the 10 μM traces being displayed in Figures 5a, b. These constructs demonstrated no loss of dimerization with truncations from the C-terminus. It further appears that the gain of secondary structure when compared to HBHA 6_199

was not due to a change in oligomeric state (Table 4). With construct 6_88 (Figure 5c), the HBHA molecule was truncated into the predicted coiled-coil region. This construct was treated the same as previous constructs and also was found to be a dimer in solution (Table 4). N-terminal truncation and leucine mutant constructs were treated as previous constructs with the 10 μM traces shown in Figure 6a (HBHA 20_199) and Figure 6b (HBHA 6_199 L14A L15A), along with the corresponding residuals. These two constructs were calculated to have the mass of a monomer (Table 4). Indeed, at all three concentrations, HBHA 20_199 and HBHA 6_199 L14A L15A, exist as a monomer in solution. Together with the CD spectra, these data indicate a link between a major loss in secondary structure signal and the conversion to monomer in solution. In contrast, constructs with alpha helical characteristics existed as dimers in solution.

Constructs were also plotted using equation 3, described in the methods. The resulting plots for each construct were linear indicating all samples are a single species in solution (Figure 7). The molecular weights of each macromolecules influence the slope of the line, resulting in the largest molecular weight construct, HBHA 1_199, having the greatest slope and the smallest molecular weight, HBHA 6_88, construct have the smallest slope. The resultant molecular weights from these plots corroborated the molecular weights achieve using the fitting software.

Crystallization Trials:

Constructs HBHA 6_199, HBHA 6_156, and HBHA 6_111 were also used to set-up crystallization trials for HBHA. Constructs were exchanged in 20 mM HEPES pH. 7.0 and 200 mM NaCl (crystallization buffer), through either dialysis or buffer exchange column. Concentrations of 10 mg/ mL for HBHA 6_199, and 20 mg/ mL for HBHA 6_156 and 6_111 were determined using the Pre-Screen (Qiagen). Trials were screened versus a myriad of sparse matrix, single component, polyethylene glycol (PEG), and pH screens (For a full list of screens

see Appendix B). None of these crystal trials resulted in crystallization of HBHA. A vector was built (see chapter 5) based upon published data to aid in crystallization. Previous work expressing a target protein fused with a previously crystallized protein was seen to yield crystals where there was previously none[28]. The cysteine protease domain of the Multifunctional Autoprocessing Repeats-in-Toxin (MARTX) protein (CPD) was chosen as a fusion partner for our target protein. The three previously mentioned HBHA constructs were expressed C-terminally fused to CPD along with a hexa-his tag C-terminal to the CPD. Expression and purification were conducted the same as previously mentioned with the exception of the TEV protease step.

Purified HBHA-fusion constructs were concentrated to previously mentioned amounts. Nickel purified samples were loaded over the Supdexdex200 size exclusion chromatography column and samples where collected. The trace of the purification of HBHA 6_111 fusion, (Figure 9a) displayed peaks of molecular weights corresponding to CPD alone and HBHA 6_111 dimer (~24 kDa). Figure 10a is an SDS-PAGE gel of the fractions (83.33 mLs and 92.84 mLs) of HBHA 6_111 + CPD untreated with phytic acid, while figure 9b has fractions from previously purified HBHA 6_111 + CPD treated with phytic acid. The bands present in both gels are similar demonstrating that the HBHA 6_111 + CPD autocleaved without the presence of phytic acid. When phytic acid is added to CPD, the N-terminal region of the protein becomes more rigid, positioning the catalytic cysteine for attack. The dimerization of HBHA 6_111 could possibly be performing the same role as phytic acid allowing the protein to cleave. To maintain the fusion, we tested multiple cysteine protease inhibitors with N-ethylmaleimide irreversibly inhibiting CPD. Each construct was treated with protease inhibitor and purified over the Superdex200 in crystallization buffer. A trace of HBHA 6_111 fusion treated with N-

ethylmaleimide (Figure 9b) displays peaks expected of the HBHA 6_111 fusion dimer. Fractions from the 75.26 mL peak, which corresponds to the fusion dimer (~71 kDa) were run on a SDS - PAGE gel (Figure 10c). All fractions show the presence of a stable fusion protein with a molecular weight of 35 kDa. Crystallization trials were setup at a range of concentrations from 10 to 20 mg/ mL of the HBHA 6_111 fusion protein.

Discussion:

Heparin Binding Hemagglutinin Adhesin is a surface protein of *Mycobacterium tuberculosis* employed for adhesion to epithelial cells; in addition, HBHA has been observed to aggregate and agglutinate in solution[29]. Recent studies have shown that HBHA exists in a dimeric state in a range of concentrations in solution. This dimer was the focus of our work. With no atomistic model available for HBHA, we employed secondary structure predictions to guide the design of constructs. These data were overlaid with the primary sequence along with the published coiled-coil prediction (Fig. 1). From the schematic of HBHA, we divided the protein into four regions: N-terminal region (blue), predicted coiled-coil (red), linker region (green) and the Heparin Sulfate binding region (orange). Constructs were designed to test the role each region may have in dimerization. Most constructs were started at residue 6, aparagine (N) versus the starting methionine. Two factors led to the decision to start at residue 6: first, the constructs that began with this N expressed well; second, they exhibited an improved half-life in solution, which was important for subsequent analytical ultracentrifugation studies. However, as shown in our data, constructs that began with either residue 1 or 6 behaved similarly throughout our biophysical characterizations.

Constructs 6_199, 6_156, 6_111, and 20_199 were initially characterized by circular dichroism spectroscopy. The trace for HBHA 6_199 appeared similar to previously published data[27]. However, as we started truncating HBHA from the C-terminus, an increase in the alpha helical signal was observed, with HBHA 6_111 being the most alpha helical (Fig. 2a). These data suggest that, *in vitro,* a majority of the residues after 111 are either disordered or in a non-alpha helical conformation, and that a vast proportion of residues 1 through 111 are in an alpha helical state.

HBHA 6_199, HBHA 6_156, and HBHA 6_111 were analyzed by AUC-SE and all were determined to have a buoyant mass of a corresponding dimer (Table 4). Unlike the previous constructs, HBHA 20_199 had a decrease in alpha helical signal and the protein adopted a more random coil conformation seen by a shifting minimum from 208 to 200 nm. HBHA 20_199 was further characterized by AUC-SE (Fig. 6a and 7) and the resultant buoyant mass was equivalent to a monomer (Table 4). These AUC-SE and CD data indicate a link between protein folding and dimerization with the N-terminal region being responsible for forming and maintaining the HBHA dimer. HBHA 6_199, HBHA 6_156, and HBHA 6_111 all contain the N-terminal region and have alpha helical CD spectra. Removal of the N-terminus causes the loss of a characteristic alpha helical signal.

Mutation of Leu-14 and Leu-15 to alanines resulted in a construct that produced a similar CD spectrum as HBHA 20_199. And, like HBHA 20_199, this construct was a monomer in solution. These data implicate the N-terminal region, in particular the predicted helix from residue 7 to 18 (Fig. 1), in dimer formation and folding. However, while the N-terminal region is important for formation of the dimer and folding of the protein, the coiled-coil is still important for the maintenance of the dimer. When residues are eliminated from the C-terminus

of the coiled-coil, there is weakening of protein stability. From the thermal denaturation experiments, HBHA 6_88 is not as stable when compared to either HBHA 6_111 or HBHA 6_156 (Fig. 3b). These observations support the conclusion that while contacts present in the coiled-coil are important for stability and preservation of the fold, it is N-terminal region that triggers folding and dimerization.

Thus, we propose that the first helix acts as a trigger sequence (Figure 8) for HBHA dimerization. Trigger sequences have been identified in other coiled-coils, where they have been shown to control folding and oligomerization[30]. HBHA is expressed predominantly on the *Mycobacterium tuberculosis* surface facing epithelial cells[15]. The heparan sulfate ligand recognized by HBHA is presented on epithelial cells in grouped brush structures[31]. Thus, we further propose that the HBHA dimer triggered by the N-terminal residues 6 - 19 of the protein serves to increase the local concentration of this adhesin such that optimal contacts with heparan sulfate brush structure can be formed. Such contacts would be expected to be critical to the initiation of pathogenesis by *Mycobacterium tuberculosis*.

Figure Legends

1. **Figure 1: Diagram of the primary sequence for HBHA aligned with a secondary structure prediction and the coiled-coil prediction.** Sequence is separated into 4 regions: blue – N-terminus, red – predicted coiled-coil, green – linker, and orange – heparin binding region. Starts and ends to constructs are shown by orange and black hashes respectively. AA represents the amino acid at the residue number above, while Sec Pred is the secondary structure predicted for that amino acid. Finally if the residue is part of the predicted coiled-coil a "c" is listed in the Coiled - Coil row.

2. **Figure 2: Far UV-CD spectra of regional HBHA constructs at 10 µM:** (A) (●) 6_199, (▲) 6_156, (♦) 6_111, and (■) 1_199. (B) (■) 6_156, (▲) 6_138, (♦) 6_111, and (●) 6_88. (C) (●) 6_199, (▼) 20_199, and (■) 6_199 L14A L15A. (D) (♦) 6_111, (▼) 20_199, and (●) 6_111 L14AL15A

3. **Figure 3: Thermal denaturation of HBHA constructs monitored at 222 nm.** (A) Constructs (●) 6_199, (■) 1_199, (▲) 6_156, and (♦) 6_111 were all at 10 µM. (B) Constructs (▲) 6_156, (■) 6_138, (♦) 6_111, and (●) 6_88 were all at 10 µM. (C) Plot of construct size (number of amino acids) versus calculated melting temperature

4. **Figure 4: Absorbance detected analytical ultracentrifugation – sedimentation equilibrium traces with residual plots.** Plot of AUC-SE for HBHA 1_199 at 3 concentrations (A) 5 µM, (B) 10 µM, and (C) 20 µM at three velocities (■)3.22 x10^4 *xg*, (▲)7.24 x10^4 *xg* and (●)1.30 x10^5 *xg* D – E: Plot of AUC-SE for HBHA 6_199 at 3 concentrations (D) 5 µM, (E) 10 µM, and (F) 20 µM at three velocities (■)3.22 x10^4 *xg*, (▲)7.24 x10^4 *xg* and (●)1.30 x10^5 *xg*

5. **Figure 5: Absorbance detected analytical ultracentrifugation – sedimentation equilibrium traces with residual plots.** HBHA constructs at 10 µM (A) HBHA 6_156, (B) HBHA 6_111, and (C) HBHA 6_88 at three velocities (■) 3.22 x10^4 *xg*, (▲) 7.24 x10^4 *xg* and (●) 1.30 x10^5 *xg*

6. **Figure 6: Absorbance detected analytical ultracentrifugation – sedimentation equilibrium traces with residual plots.** HBHA constructs at 10 µM (A) HBHA 20_199 and (B) HBHA 6_199 L14A L15A at three velocities (■) 3.22 x10^4 *xg*, (▲)7.24 x10^4 *xg* and (●)1.30 x10^5 *xg*

7. **Figure 7: 7.24 x10^4 *xg* AUC Data of HBHA constructs plotted as a relation of molecular weight.** (●) HBHA 1_199, (■) HBHA 6_199, (▲) HBHA 6_156, (▼) HBHA 6_111, (♦) HBHA 6_88, (●) HBHA 6_199 L14A L15A, and (.) HBHA 20_199

8. **Figure 8: Schematic of HBHA Model.**

9. **Figure 9: Size exclusion chromatography traces of HBHA 6_111.** (A) Purification of HBHA 6_111 + CPD without N – ethylmaleimide (B) Purification of HBHA 6_111 with N – ethylmaleimide

10. **Figure 10: SDS -PAGE gel of HBHA 6_111 + CPD.** (A) lane 1 fermentas ladder, lane 2 – 83.33 mLs peak, lane 3 – 92.84 mLs peak. (B) lane 1 – phytic acid treated sample, lane 2 – imidazole eluted fraction, lane 3 – Bio-Rad unstained ladder. (C) Lane 1 – Bio-Rad unstained ladder, Lanes (2-9) fractions A8 – B10.

Chapter 3:

Function of Mt Manb: A Mycobacterial Phosphomannomutase

Introduction

The pathogenic pathway of *Mycobacterium tuberculosis* involves invading and parasitizing the host alveolar macrophages. One possible interaction between the macrophage and mycobacteria involves the C-type lectin mannose receptor and the mannose-capped lipoarabinomannan (ManLam) molecules on the respective cell surfaces[32]. The C-type lectin mannose receptor (MR) recognizes carbohydrates on the surface of microorganisms[33]. Binding via the MR triggers engulfment without necessarily causing a pro-inflammatory response[34]. Phagocytosis, via the C-lectin receptor, prevents phagosome maturation resulting from the inhibition of calcium ion production in the macrophage cytosol[35]. Downstream processes that recruit early endosome antigen 1 to the phagosome are inhibited preventing the fusion of the phagosome to the endosomal-lysosomal vesicles allowing the mycobacterium to exist within the vacuole[32].

The ManLam molecule, which binds to the Mannose receptor, is built off a lipomannan backbone (Figure 11). The lipomannan (LM) chain is composed of an alpha 1→6 Mannose phosphate scaffold anchored to phosphatidlyinositol backbone. The further addition of arabinose branch structures extends the complex molecule further away from the cell surface. Finally the lipoarabinomannan (LAM) structure is capped by mannose polysaccharides. Synthesis of the

nature of ManLam structures are still being investigated, but the mannose donor molecule is exclusively guanosine diphosphate mannose (GDP-mannose)[36]. The synthesis of the GDP-Mannose molecule happens through a series of enzymatic reactions (Figure 12). Mannose – 6 - phosphate enters the pathway through two possible paths. The first involves mannose uptake from the extracellular environment and conversion to mannose – 6 – phosphate (M6P) via a hexokinase. The second path requires the conversion of glucose-1-phosphate to mannose-6-phosphate by means of several enzymes, including isomerases and mutases[37]. The mannose - 6 - phosphate is then converted into mannose – 1 – phosphate (M1P) via a phosphomannomutase. Mannose-1-P is further process using the enzyme GDP-mannose pyrophosphorylase generating the donor molecule, GDP-mannose. Both pathways converge at the mannose 6 phosphate to mannose 1 phosphate, which is facilitated by phosphomannomutase, MtManb (Rv3257c)[37].

Background

Mt. Manb is a member of the phosphohexomutases, which interconvert M6P to M1P or glucose – 6 – phosphate to glucose – 1- phosphate. Members of the phosphohexomutase family use catalytic serine residues to perform their function and require magnesium for catalysis to occur[38]. Prior to substrate binding, the catalytic serine is phosphorylated by an unknown mechanism. In vitro experiments, a 1,6 - bis-phosphate hexose sugar is used to "prime" the reaction[39]. Upon substrate binding (e.g. M6P) the phosphorylated serine donates the phosphate group generating a 1,6 - bis-phosphate sugar intermediate. The intermediate sugar is then rotated 180 degrees within the binding pocket reorienting the 6 position phosphate near the catalytic serine[39]. The serine residue removes the 6 – phosphate from the sugar intermediate leaving only the 1 – phosphate. The product is released and the catalytic serine is again phosphorylated and

ready to perform another cycle. The focus of this work was to obtain the x-ray crystal structure of Mtmanb along with designing a high throughput assay to test potential inhibitors. Inhibition of Mtmanb would prevent synthesis of ManLam molecules on the surface of pathogenic mycobacterium preventing macrophage phagocytosis via the mannose receptors, which would subsequently allow for phagosome – lysosome maturation and bacterial death.

The current activity assay for phosphohexomutases involves coupling reactions with several enzymes (Figure 13) where the ultimate signal is the change in $O.D._{340}$ due to the conversion of $NADP^+$ to NADPH. M1P is converted to M6P by phosphomannomutase followed by the conversion to fructose – 6 – phosphate (F6P) by phosphomannose isomerase. F6P is then converted to G6P by phosphoglucose isomerase, and finally G6P is converted to 6-phosphogluconate while $NADP^+$ is reduced to NADPH. Due to potential inhibition of ancillary enzymes, this assay is not optimized for screening inhibitor libraries via high throughput methods. Each protein in the coupled assay would have to be screened individually for inhibition. A new assay we are proposing has one coupling enzyme, Mannose-6-P reductase, which uses NADPH to reduce M6P, and in the process it, NADPH, is converted to $NADP^+$. Like the established assay, the change in OD_{340} is being measured. The reduction in coupling proteins should be a benefit to high throughput screening.

Construct Design and Protein Expression

The gene for Mtmanb (Rv3257c) was cloned out of H37Rv genomic DNA and inserted into pUC19 using polymerase chain reaction and ligation reactions. Based on previously crystallized homologs listed in the Protein Data Bank (PDB)[40] and secondary structure predictions aligned with domain function, the full length protein was the only construct designed

and cloned into the bacterial expression vector pMCSG7[21]. Cloning into the expression vector was discussed in chapter 2 and sequence verified genes were transformed into BL21 (DE3) cells and BL21 (DE3) – RIPL cells due to the presence of rare codon clustering in the DNA sequence. Test expression studies were performed using TB media for overnight and LB media in 4 hour inductions. Cells were lysed and protein expression was verified via SDS-PAGE gel. Protein was purified over nickel sepharose column (GE healthcare) and his-tag was removed using TEV protease overnight dialyzing against TEV buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM DTT) at 10^{o}C. The cleavage reaction was re-passed over a nickel column removing the histidine tag and any uncleaved material. Cleaved sample was collected and concentrated to a volume < 5mL. Concentrated sample was injected over a Superdex200 size exclusion column (GE Healthcare) (Figure 14). Protein from the 81 mL peak were combined and used for crystallization trials in 20 mM Tris – HCl pH 8.0 100 mM NaCl.

Mannose – 6 – Phosphate Reductase

The Arabidopsis thaliana gene for mannose – 6 – phosphate redutase (M6PR) was ordered from the Arabidopsis Biological Resource Center (ABRC) at Ohio State. The gene came inserted into pUC18 and transformed into DH5. Using sterile technique, an innoculum loop was used to streak the bacteria on a LB + Ampicillin plate. The plate was grown at 37^{o}C was grown overnight. Colonies from were used to inoculate 6 mL LB + Amp cultures, grown overnight. DNA was harvested using Qiagen® miniprep kit. The full length M6PR gene was cloned out of pUC18 using PCR. The gene was then inserted into pMCSG7 using LIC method. Sequence verified vectors were transformed in Bl21 (DE3) competent cells for expression. 1 Liter cultures were grown to an O.D.$_{600}$ of 0.8. A final concentration of 0.75 mM IPTG was used

to induce expression of M6PR. Cells were harvested after 4 hrs and were pelleted at 4500 *xg*. M6PR was purified following the same protocol used for His-Lic Mtmanb.

Discussion

Rv3257c is a phosphomannomutase used in the production of GDP-mannose donor molecule for synthesis of the ManLam structures on the surface of *Mycobacterium tuberculosis*. The interaction of the ManLam with mannose receptors on macrophages has been shown to inhibit phagosome – lysosome maturation[37]. Inhibition of ManLam synthesis would allow the macrophages to ingest the bacterium via another receptor allowing for proper disposal of the invading microbe. Our goal for the project was to get the three dimensional crystal structure of Rv3257c and develop a high throughput method for screening inhibitors.

From previously crystallized homologs (PDB #: 2FKF, 1P5D, and 1K2Y) and secondary structure prediction the full length construct was chosen as the target. The classical PGM_PMM domains extend from residue 9 to residue 464. Prior to setting up Mtmanb crystals trials, a pre-screen technique was used to determine a concentration range for the trials[41]. Our concentration range for crystallization was 10 mg/ mL to 17 mg/ mL. Previously crystallized homologs were crystallized at this concentration range[38]. Crystallization trials were setup against screen listed in Appendix B. None of these conditions yielded diffraction quality crystals. Conditions similar to those setup by Regni *et al.*[38] were also tried with no crystal formation.

When comparing Mtmanb sequence to 2FKF the sequences were 35.6 % identical and a 52% homology. Using the PHYRE server a three dimensional model was generated for Mtmanb using it homologs as references[42]. The three dimensional model from Mtmanb and three dimensional structure of 2FKF can be seen in figure 16 a,b, and c[40]. The alignment of the 2FKF

structure and the Mtmanb model, shows similarities in overall fold. Next we generated the surface charge potentials of the 2FKF structure and the Mtmanb model (Figure 17 a, b) using PyMOL software[43]. Surface charge potentials are based on amino acids present in the protein at particular area where regions of negative charge are colored red and regions of positive charge are blue. The mask generated from the surface charge potentials are based on Poisson - Boltzman calculations and are meant to be a generalized representation of the charged surface of the protein. The charge masks help demonstrate the differences between the two proteins, even with high homology and identity. The surface charge difference may have led to failure of crystallization. In figure 17 a, b, the two surface masks have matched orientation. The regions of the protein surround by the yellow box mark the surface where symmetry mates interact (crystal contacts). The insufficient formation of intermolecular contacts is one of the main barriers to stable lattice formation[44]. In figure 17b, 2FKF symmetry mate contacts a region of positive charge, while the same region is negatively charged in the model of Mtmanb. If we rotate each structure clockwise 90^{o} (Figure 18), another region of symmetry mate contact is highlighted by the yellow box in figure 18b. Again there is a difference in charge present at the site of contact. Two regions of possible intermolecular contact between symmetry mates are possibly not present in a Mtmanb lattice. The difference in surface charge distribution might be one of the reasons for a breakdown in the crystallization of Mtmanb.

Figure Legends

11. **Figure 11: Structure and Biosynthesis of ManLam**

12. **Figure 12: GDP-mannose Synthesis Pathway**

13. **Figure 13: Phosphohexomutase Coupled Assay Schematic**

14. **Figure 14: SEC200 Purification of Rv3257c.** Aliquots were taken during each step in Rv3257c purification. Lane 1 – Protein ladder, 2 lysate, 3 flow through, 4 Wash, 5 elution, 6 – 7 TEV fraction elution, 8 TEV fraction pre flow through, 9 flow through TEV fraction (used for crystallization)

15. **Figure 15: Ribbon Diagrams of Mt Man Model and PDB Structure 2FKF.**

16. **Figure 16: Charge potentials diagrams from Mtmanb model and 2FkF from the PDB**

17. **Figure 17: Charge potential diagrams from Mtmanb model and 2fkf rotated 90 degrees about the z-axis.**

Chapter 4:

Two Component Sensor Histidine Kinase (DevS)

Introduction

The treatment for Tuberculosis is multi-month regimen of antimicrobials that if improperly administered allows the mycobacterium to become resistant the treatment. Maintaining the regimen in poor or developing countries takes great effort due to limited access to physicians to monitor patient's progress and money for treatment. The cause for an extended therapy is due to *Mycobacterium tuberculsosis'* ability to undergo metabolic transformation to a dormant state or in a state of non-replicating persistence (NRP) within granuloma's[45]. Current antibiotic treatment is only effective against active mycobacteria not bacteria in NRP, so a constant level of antimicrobials needs to be present to ensure complete mycobacterial death. NRP has been shown to be triggered by hypoxia or the presence of nitric oxide (NO), which is produced in macrophages via inducible Nitric Oxide Synthase (iNOS)[46]. NRP is induced by up-regulating 48 genes under the control of the DevSR two-component system[47].

Two Component System

Two component systems are composed of a histidine kinase (HK) and a response regulator (RR) used to process the signal in response to some external stimuli. DevS is a sensor histidine kinase that reacts to the change in the macrophage environment. Once the appropriate conditions are present, the protein undergoes autophosphorylation at His 395[45]. The phosphate is

transferred to aspartate 54 (Asp 54) of the RR, DevR, using Mg^{2+} as a co-factor[48]. DevR then triggers up-regulation of dormancy genes, guarding the mycobacterium from the macrophage/ granuloma environment along with protecting the bacterium from Tuberculosis treatment[49]. To increase the efficacy of the current treatment, the addition of an anti-dormancy compound should be included in the treatment regime.

There are approximately 11 different two component systems characterized in *Mycobacterium tuberculosis*, whose functions range from chemotaxis to osmoregulation [50]. In a comparison of avirulent and virulent strains of *Mycobacterium tuberculosis*, the DevS – DevR system (differentially expressed in the virulent strain) was the shown to be up-regulated in virulent strains during infection[50]. Recent work by *Voskuil et al*., suggested that the redox state is the triggering event and they demonstrate a path by which the respiratory state of the bacteria can be linked to DevS activation [51]. The focus of this work was to the sensor histidine kinase, DevS.

Background

DevS is a 578 amino acid long protein compartmentalized into four distinct regions, starting from the N-terminus; GAF (c**G**MP-specific phosphodiesterases, **a**denylyl cyclases and **F**hlA) A (residues 63 to 200), GAF B (residues 231 to 369), histidine kinase (residues 386 to 452), and HSP (heat shock protein) - 90-like ATPase (residues 488 to 577). These domains work in concert to process the dormancy signal from outside stimuli to expression of proteins such as alpha crystallin[52]. The GAF domains (A and B) are commonly found in tandem in most proteins, where one function's as a small molecule regulatory domain and the other can either have similar function or act as a dimerization domain[53]. The most common cofactors bound by these domains

are cyclic nucleotides and no other GAF domains have been shown to bind heme[53]. In contrast the PAS (period circadian, aryl hydrocarbon receptor nuclear translocator protein, single-minded protein) domains bind to heme as a co-factor[53].

From work done by Sardiwal *et al.*, it was determined that the first GAF domain, GAF A, binds to heme while the function of the second GAF domain, GAF B is undetermined[54]. GAF domains and PAS domains are structurally similar, but the sequences are unrelated and each domain binds heme in different orientations[55]. GAF A binds the beta-Heme perpendicular to the beta-strands while PAS domains bind heme parallel to the beta-strands (Figure 19 a, b)[55]. It has been shown that histidine 149 (H149) is important in heme binding in the GAF A domain. When the residue is mutated to alanine (H149A) the domain is no longer capable of binding heme[54].

NO binding to the iron center or the absence of diatomic oxygen, triggers autophosphorylation[56]. *Yukl et al.* describe tyrosine 171 (Y171), hydrogen bonding to the bound O_2 molecule, bound to the iron, being the inactive form of the molecule[56]. Upon NO binding or O_2 displacement, Y171 moves out of position and potentially interacts with E87 as part of a hydrogen binding network[55]. Mutation of Y171 to phenylalanine did not cause the protein to become constitutively active, but was incapable of activity, demonstrating the relationship between Y171 and activity[55]. The crystal structure of GAF A domain has been solved in several redox states and bound ligands. In none of these structures were there any large conformational changes that would hint to the mechanism of autophosphorylation. The authors thought this might be due to the constraints of crystal packing[55].

The structure of the GAF B domain from *Mycobacterium smegmatis* DevS has also been solved, but functional questions still remain[57]. Other members of the GAF superfamily use the

second GAF domain as either a small molecule binding pocket or dimerization domain, but none of these functions are present in GAF B[57]. GAF B does not bind to any tested ligand, nor has DevS been seen as a dimer in vitro. Some work has shown that GAF B increases the specificity of O_2 bound to heme, as the stabilizing molecule for the hydrogen bonding network [58]. How this signal is transferred to the kinase domain is still unknown and it is possible that the GAF B domain acts as a transducer for the signal from GAF A to the histidine kinase/ ATPase regions of DevS.

The C-terminal region of DevS consist of the histidine kinase (HK) and ATPase domain. The donor phosphate comes from the hydrolysis of ATP (adenosine-5'-triphosphate) by the ATPase (residues 488 to 577) relieving the gamma phosphate. The phosphate is then passed to the RR, DevR, causing up regulation of the dormancy regulon. The focus of this work was to crystallize the four domains of DEVS in both the active and inactive forms to elucidate how hypoxia or NO binding triggers auto-phosphorylation and subsequently dormancy.

Construct Design, Protein Expression and Purification

Designing constructs for DevS proceeded as mentioned in chapter 2 with initial analysis of secondary structure and domain organization. The gene for DevS (Rv3132c) was cloned from *M. tuberculsosis* H37Rv genomic DNA using PCR amplification into pUC19 replication vector. The primary sequence was input into the CCD - NKI site for secondary structure analysis (Figure 20)[20]. Results from these data were used in designing constructs for protein expression. We chose to express three constructs: Full length, 15_578, and 62_578. Crystal structures of the GAF A and GAF B have been solved independently of each other[53,57]. Crystallization of both GAF domains together along with the HK and ATPase domains are important to understanding

how the transmission of signal is accomplished. All constructs were terminated at residue 578 on the C-terminus. The N-terminal start of our constructs varied. Along with secondary structure predictions, software was used to predict whether there were transmembrane regions of the protein. From previous literature, there it was hypothesized that transmembrane helices were present allowing for the sensor region to have access to the signal [50]. Constructs including these regions have been expressed using an *E.coli* expression system, so they were included in our designed constructs[50]. No signal peptide was discovered using SignalP 3.0 software, so the full length construct was included in the list[59]. Since it is likely that there are signal sequences for *Mycobacteria spp.* that are not discovered, two other constructs were designed starting at methionine 15 (M15) and the third start is at glycine 61 (G61) proximal to the start of the GAF A domain. Upon further analysis of the codon usage, we determined that codon optimization might help with increasing expression. Protein was submitted to GENSCRIPT™ for codon optimization.

Each DevS construct was inserted into three expression vectors. The vectors were pMCSG7, pET-32 EK/LIC (EMD Biosciences), and pCPD Lasso (Appendix C). Each expression vector contained T7 promoters, IPTG inducible, and a his tag (hexa- or deca-). Vector pMCSG7 was described in previous chapters. The pET32 EK/LIC vector fuses a thioredoxin protein N-terminal to the target protein connected to the via an enterokinase cleavage site to remove the tag. The pCPD Lasso vector attaches the CPD fusion protein C-terminal to the target protein and is connected to the deca-his tag via a ten residue linker of serine and glycine residues. Sequence verified sequences were transformed into BL21 (DE3) GOLD cells along with L-arabinose inducible GroEL/ ES. Expression of DevS constructs were performed as described by *Yukl et al.* [56]. Cells were grown to an $O.D._{600}$ of 0.8 when then temperature was

lowered to 18^{o}C. An L-arabinose solution was added until a final concentration of 0.2% v/v. Cells were allowed to shake for 30 minutes. IPTG was then added to the cells at a final concentration of 0.5 mM and left to grow overnight. Cells were pelleted and resuspended in 50 mL of lysis buffer (100 mM Potassium phosphate pH 7.4, 500 mM NaCl, 10% glycerol, 0.5 mM TCEP) per 2 Liters of culture. Resuspended pellets were aliquoted into 50mL culture tubes then stored at -80^{o}C.

Cells were lysed in 50 mL of lysis buffer plus protease inhibitors, lysozyme, DNase, 5 mM MgCl and 5 mM ATP. MgCl and ATP were added to lysis mixture to help with any possible chaperone attachment to the target protein. Solution was allowed to equilibrate for 30 minutes. Cells were lysed using a Branson Sonic dismembraner with 3 cycles on .5 sec pulse/ 1 sec rest at 30% power. Cell lysate turned a rose color due to the oxidation of the heme present in solution. Lysed cells were centrifuged at 16 K RPM for 70 minutes at 10^{o}C. Samples were decanted collecting the supernatant. Supernatant was passed through a 0.22 um syringe filter. Syringe filtered sample was loaded over a HisTrap HP column (GE Healthcare) collecting flow through. Column was washed with lysis buffer until a protein signal could no longer be detected by Bradford assay or UV detection when using the AKTA Express® (GE Healthcare). The column was then washed with 30 mM imidazole in lysis buffer, removing non-specifically bound material. Protein was eluted from the nickel column using a 500 mM imidazole plus lysis buffer bump and fractions were collected. The following step is where the purification procedure diverges for each expression system; LIC system, EK/LIC and CPD system. Using the LIC system the protein was put into a dialysis membrane (7500 MWCO) with the addition of 0.5 mg of TEV protease. Sample was dialyzed against 50 mM Tris pH 8.0, 100 mM NaCL, and 1mM DTT and left to go overnight at 4^{o}C. The EK/LIC was treated similarly to the TEV protocol, but

enterokinase was added instead. Using the CPD system, a solution of inositol-6-phosphate (Ins6p) was added at a final concentration of 1 mM to elicit auto-cleavage. Sample was allowed to gently shake overnight at 4^{o}C. Each reaction was again passed over a nickel column re-equilibrated in lysis buffer, no imidazole, collecting flow-through. The rose color was still present in the flow through. A SDS - PAGE gel was run on lysate, flow through, elution, cleavage reaction, and second flow through fractions. Gel was analyzed to see the presence of target protein and also to see if cleavage worked (Figure X.X). Samples from the second nickel column were concentrated using a centrifugal concentrator MWCO of 30 kDa to a volume of less than 5 mL. As the sample was concentrated the color of the sample increased in redness, indication that heme was bound to protein. The concentrated sample was loaded over a Superdex S200 SEC that was pre-equilibrated in 10 mM Tris pH 8.0, 100 mM NaCl, and 0.5 mM TCEP. Peak fraction were collected and run on a SDS - PAGE gel to verify target protein. Fractions from the SEC 200 purification lacked the reddish color that was present throughout purification indicating a loss of bound heme. Elution volume off the column also indicated that the protein was soluble aggregate.

Discussion

Recombinant expression of DevS requires the addition of heme in the form of the ferric salt hemin (sigma-aldrich) along with charperones GroEL/ES[58]. Omission of these cofactors contributes to a decrease in DevS production. In our experiments we had no problems with the amount of expression, but misfolded expression

Signs of misfolded protein presented as a doublet around the molecular weight of DevS on a SDS – PAGE gel. The lower band always co-eluted with DevS. The molecular weight of

the band is similar to that of the calculated value for GroEL suggesting that DevS is misfolded and stuck within the chaperone. One of the methods used to remove stuck chaperones is the addition of ATP to the lysis buffer. It is hypothesized that the chaperones will turnover the ATP and cycle through until the target is clear of the complex. During the lysis process, ATP was added to lysis buffer along with $MgCl_2$ (added to protect the nickel column), but the secondary band still co-eluted. Recombinant expression of full length DevS and DevS 62_578 has been achieved at low levels of expression, but never at high enough concentrations for crystallization.

One possible method to overcome misfolding is to decrease the temperature during growth to slow down the production and folding of the recombinant protein. DevS expression, full length and DevS 62_578, were repeated using ArticExpress® cells in place of BL21 (DE3). GroEL/ES chaperones at low temperatures (<16^{o}C) have a decreased activity[60]. ArticExpress® cells have endogenous chaperones that are 100% active at temperatures between 4 and 12^{o}C.

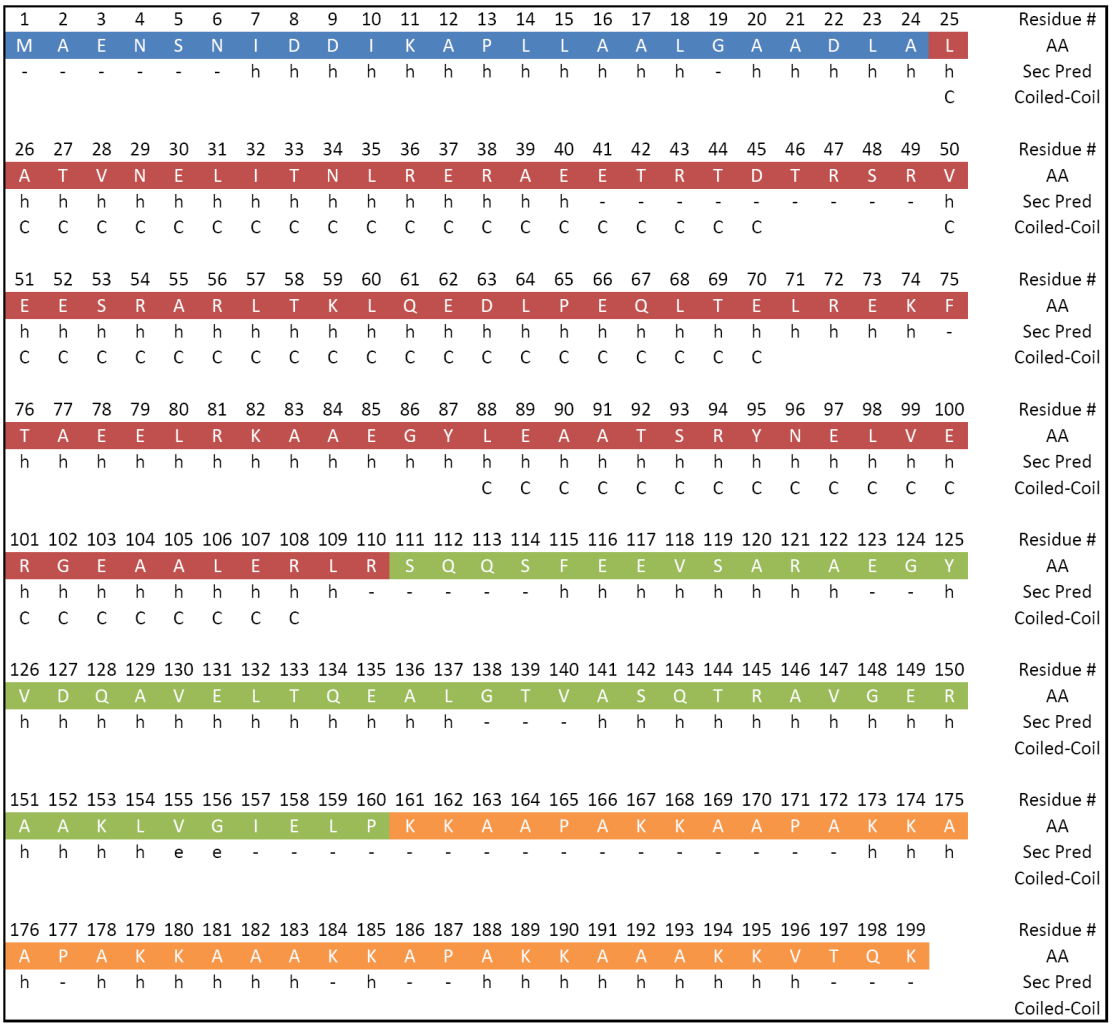
Recombinant DevS in the EK/ LIC TRX expression vector was transformed in ArticExpress® cells. Method of expression was performed similarly as previously described by Yukl *et al.*, except with changes to the cell type, the temperature was lower to 12^{o}C and the induction was left to run for 36 hrs. Cells were lysed at following prior protocols. Samples were run on a SDS – PAGE gel. Expression of DevS 62_578 seemed to work and we were able to separate the doublet band (Figure 20).

Purified DevS was concentrated using spin concentrator with a 30K MWCO to a volume of less than 5 mL. Heme was still detected in solution due to the presence of the rose color. Sample is loaded over sizing column that has been equilibrated in 10 mM HEPES pH 8.0, 200 mM NaCl, 1 mM TCEP, and 2% glycerol. Peak fractions were collected based on

chromatogram (Figure 21) and fractions were run on a SDS – PAGE gel. Collected fraction no longer was a rose color. During purification over the S200 sizing column the heme was removed from DevS. The Heme co-factor was consistently lost after purification off the sizing column. After loss of heme, DevS would aggregate.

Figure Legends

Figures



Residue #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
AA	M	A	E	N	S	N	I	D	D	I	K	A	P	L	L	A	A	L	G	A	A	D	L	A	L
Sec Pred	-	-	-	-	-	-	h	h	h	h	h	h	h	h	h	h	h	h	-	h	h	h	h	h	h
Coiled-Coil																									C

Residue #	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
AA	A	T	V	N	E	L	I	T	N	L	R	E	R	A	E	E	T	R	T	D	T	R	S	R	V
Sec Pred	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	-	-	-	-	-	-	-	-	-	h
Coiled-Coil	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C					C

Residue #	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
AA	E	E	S	R	A	R	L	T	K	L	Q	E	D	L	P	E	Q	L	T	E	L	R	E	K	F
Sec Pred	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	-
Coiled-Coil	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C					

Residue #	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
AA	T	A	E	E	L	R	K	A	A	E	G	Y	L	E	A	A	T	S	R	Y	N	E	L	V	E
Sec Pred	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h	h
Coiled-Coil													C	C	C	C	C	C	C	C	C	C	C	C	C

Residue #	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125
AA	R	G	E	A	A	L	E	R	L	R	S	Q	Q	S	F	E	E	V	S	A	R	A	E	G	Y
Sec Pred	h	h	h	h	h	h	h	h	h	-	-	-	-	-	h	h	h	h	h	h	h	h	-	-	h
Coiled-Coil	C	C	C	C	C	C	C	C																	

Residue #	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
AA	V	D	Q	A	V	E	L	T	Q	E	A	L	G	T	V	A	S	Q	T	R	A	V	G	E	R
Sec Pred	h	h	h	h	h	h	h	h	h	h	h	h	-	-	-	h	h	h	h	h	h	h	h	h	h
Coiled-Coil																									

Residue #	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
AA	A	A	K	L	V	G	I	E	L	P	K	K	A	A	P	A	K	K	A	A	P	A	K	K	A
Sec Pred	h	h	h	h	e	e	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	h	h	h
Coiled-Coil																									

Residue #	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199
AA	A	P	A	K	K	A	A	A	K	K	A	P	A	K	K	A	A	A	K	K	V	T	Q	K
Sec Pred	h	-	h	h	h	h	h	h	-	h	-	-	h	h	h	h	h	h	h	h	h	-	-	-
Coiled-Coil																								

Figure 1: Diagram of the primary sequence for HBHA aligned with a secondary structure prediction and the coiled-coil prediction.

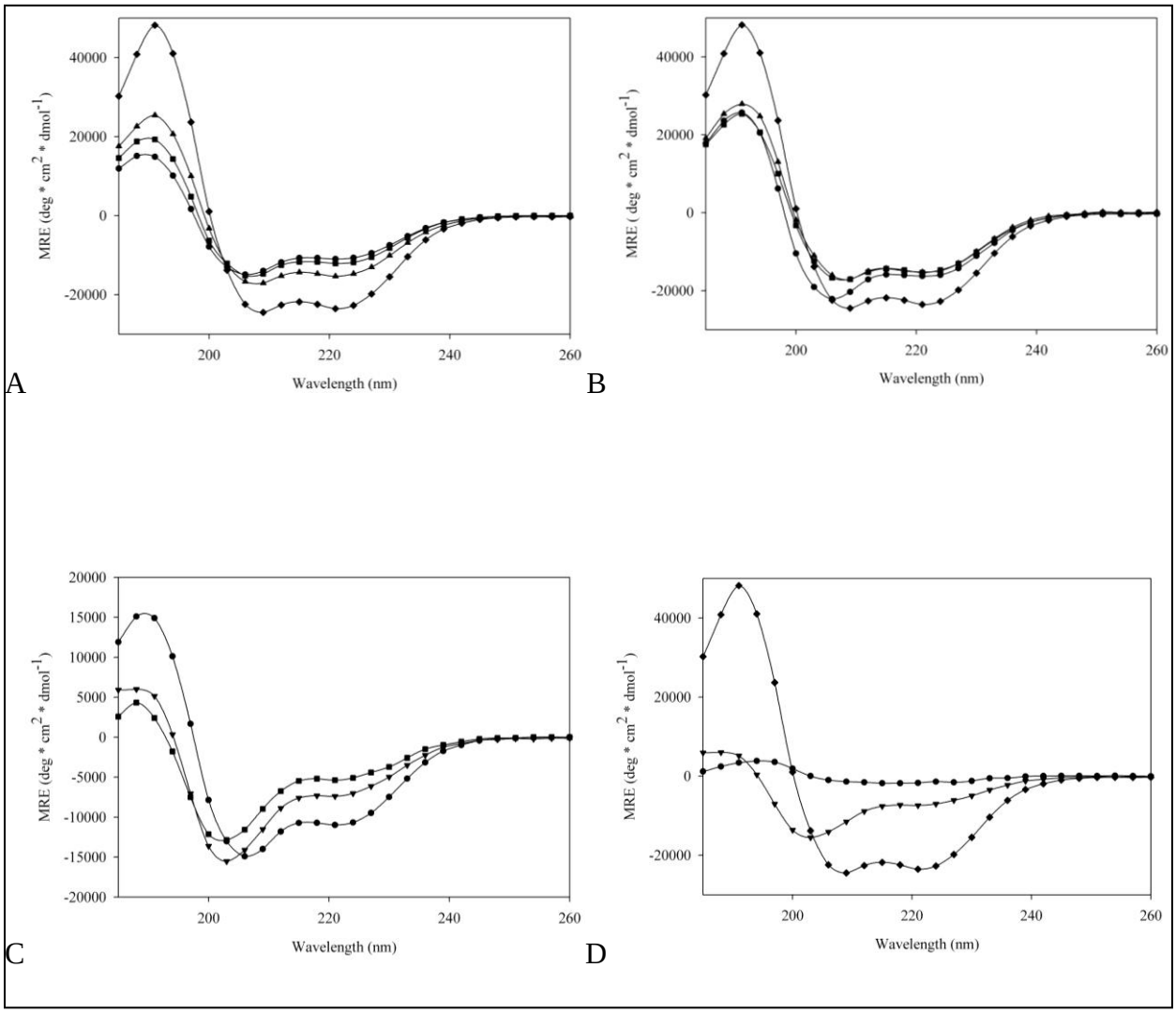


Figure 2: Far UV-CD spectra of regional HBHA constructs at 10 µM

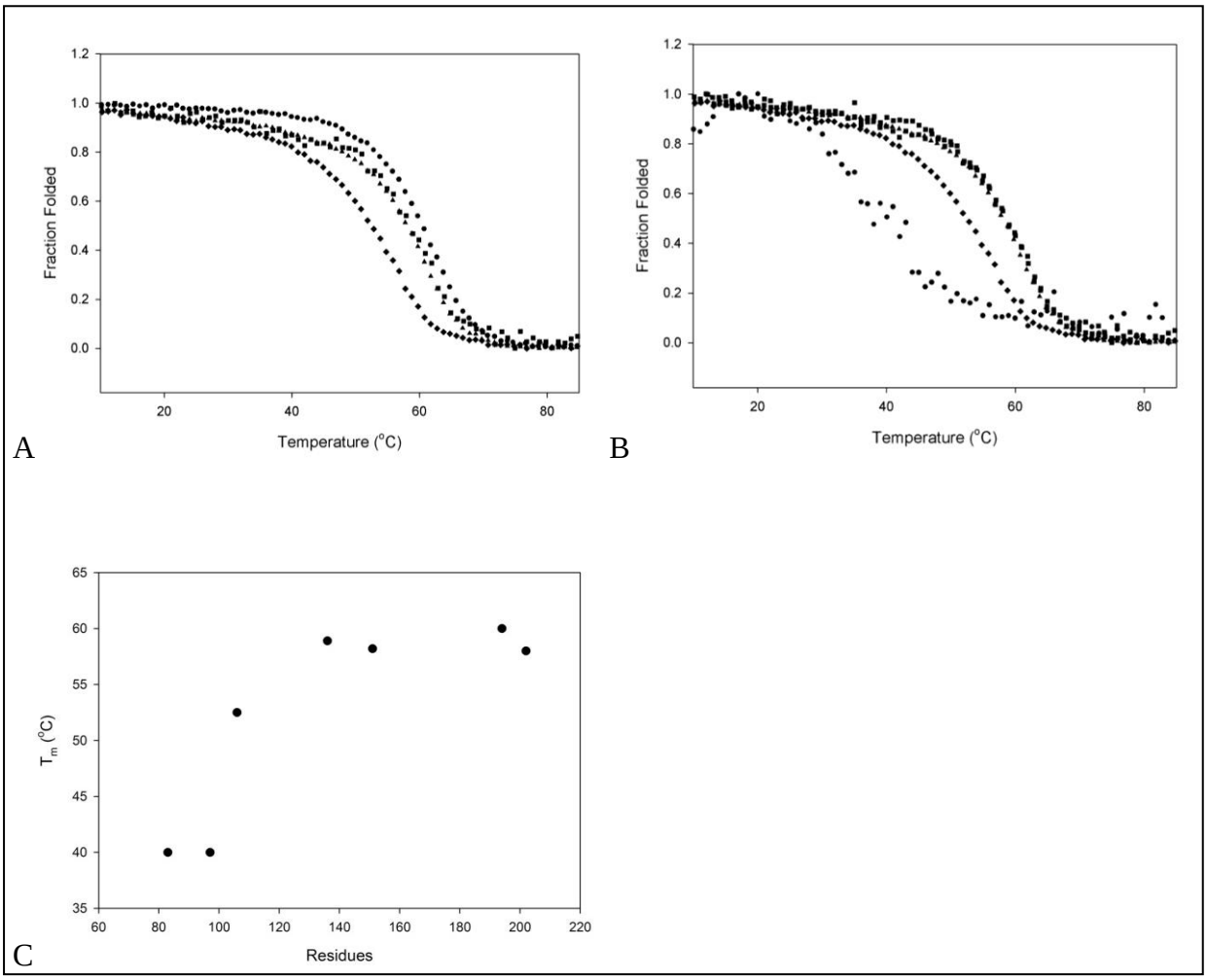


Figure 3: Thermal denaturation of HBHA constructs monitored at 222 nm.

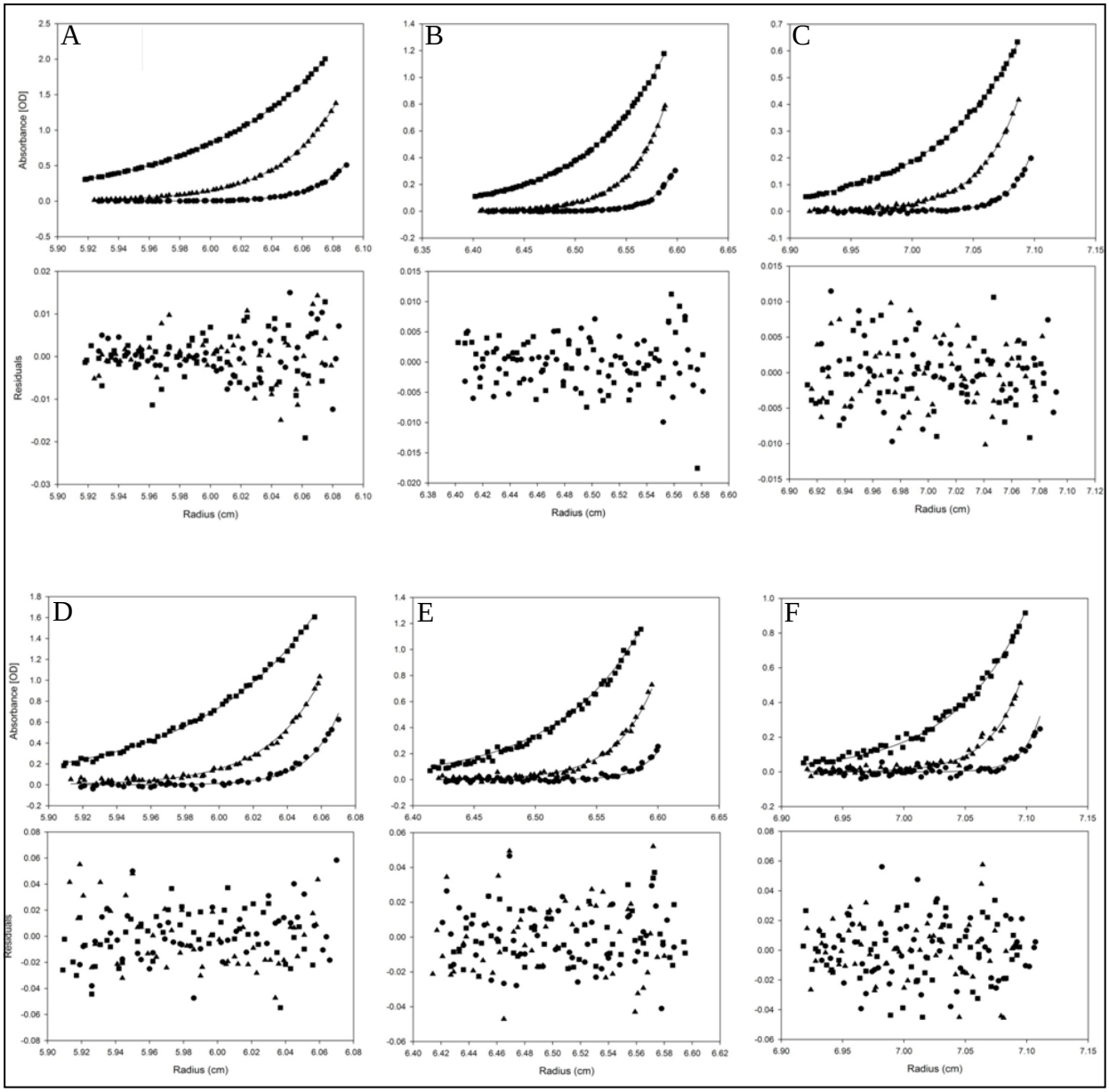


Figure 4: Absorbance detected analytical ultracentrifugation – sedimentation equilibrium traces with residual plots.

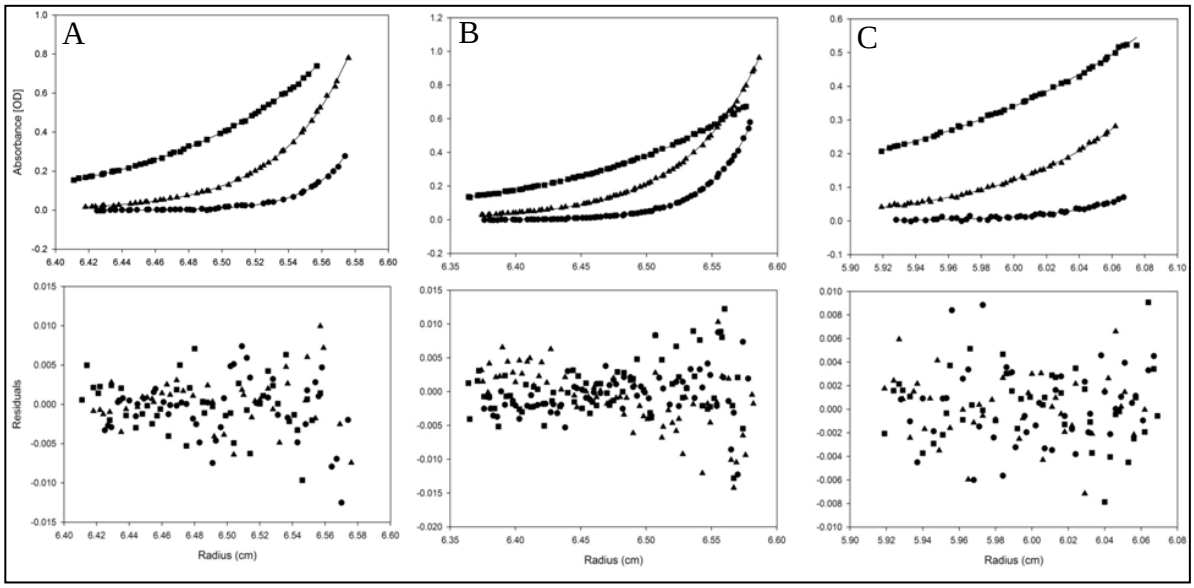


Figure 5: Absorbance detected analytical ultracentrifugation – sedimentation equilibrium traces with residual plots.

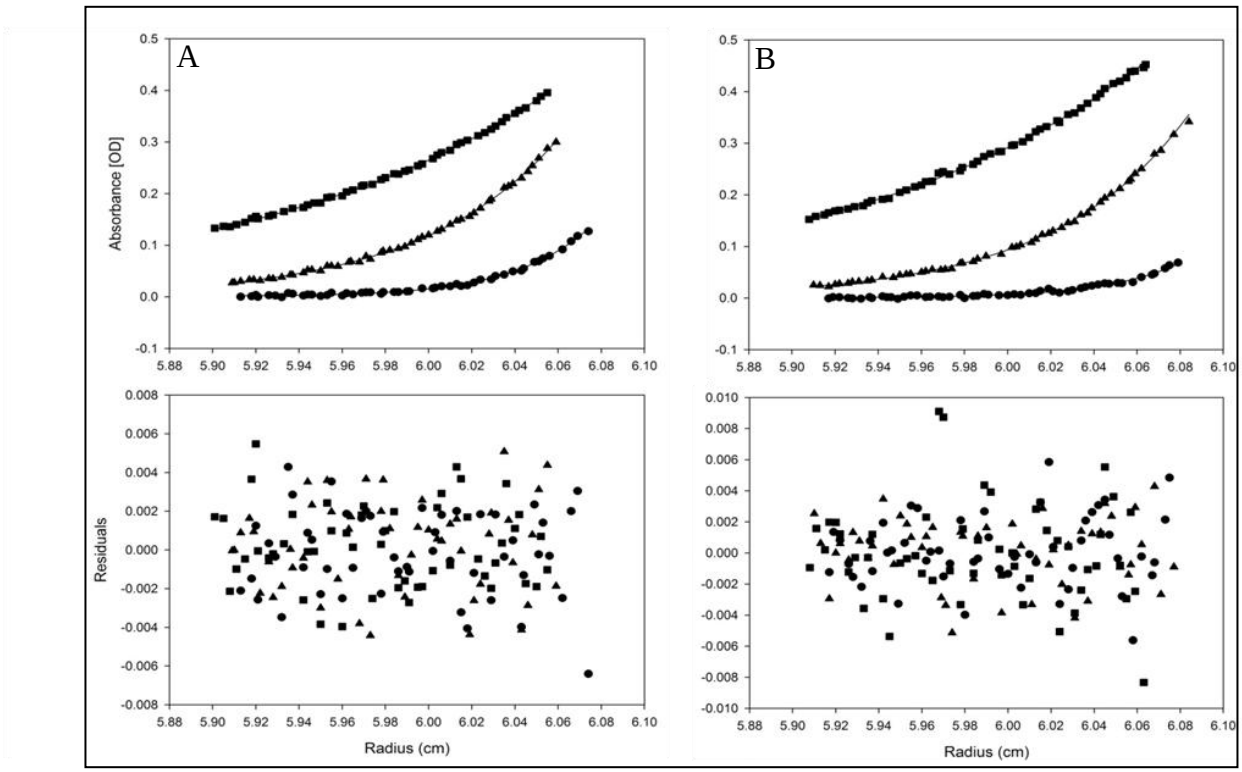


Figure 6: Absorbance detected analytical ultracentrifugation – sedimentation equilibrium traces with residual plots.

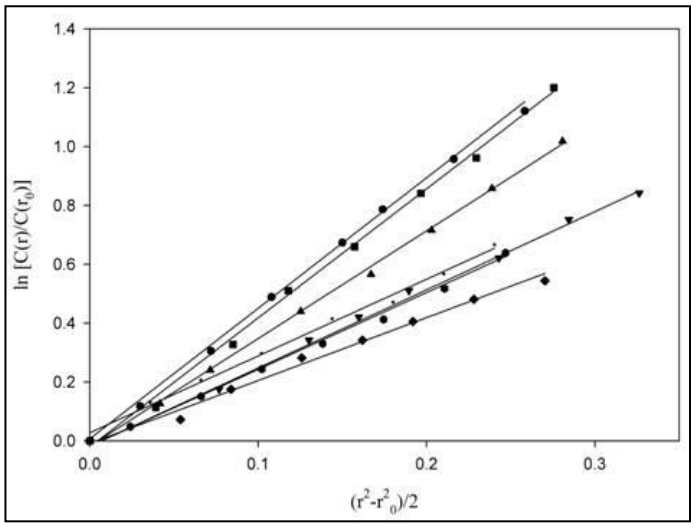


Figure 7: 7.24 x10^4 *xg* AUC Data of HBHA constructs plotted as a relation of molecular weight.

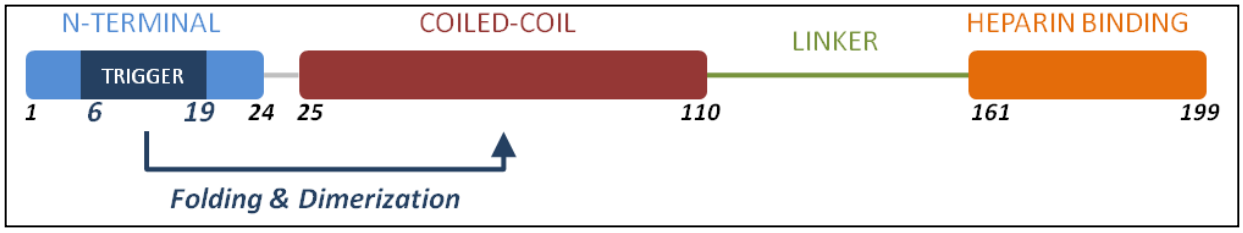


Figure 8: Schematic of HBHA Model.

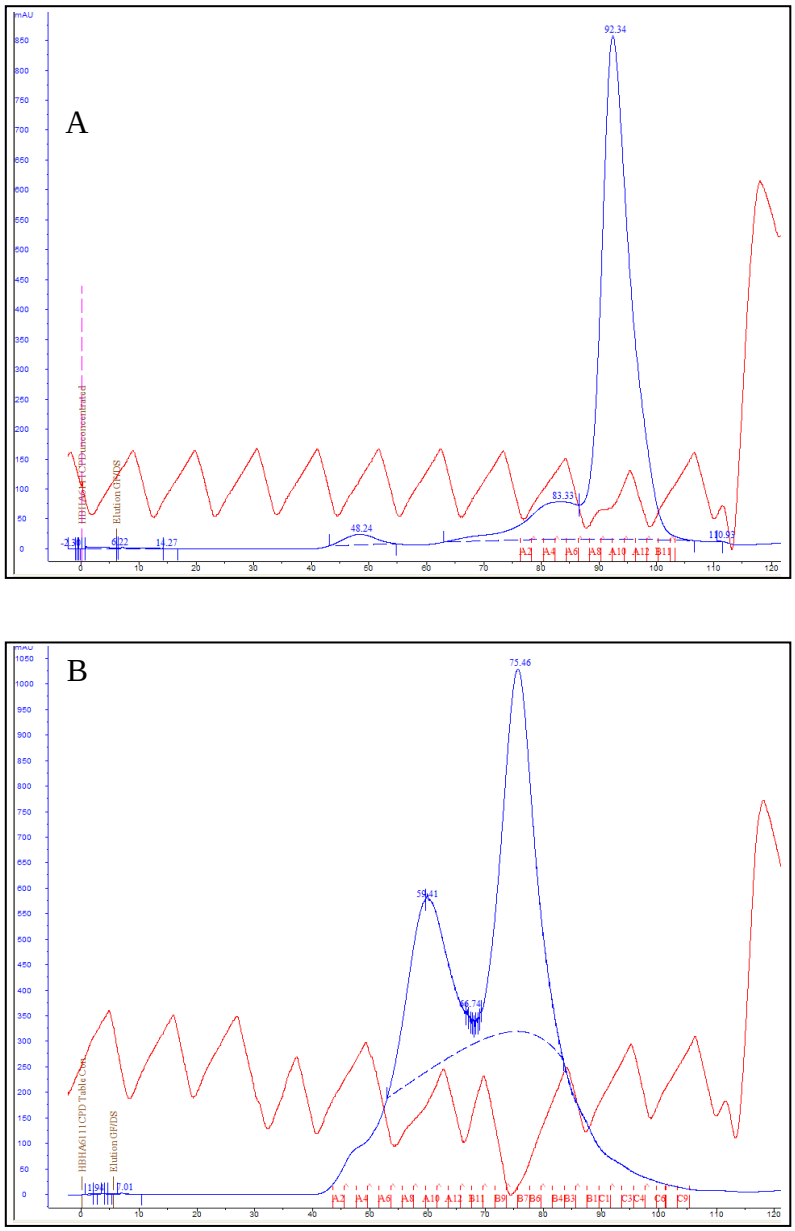


Figure 9: Size exclusion chromatography traces of HBHA 6_111.

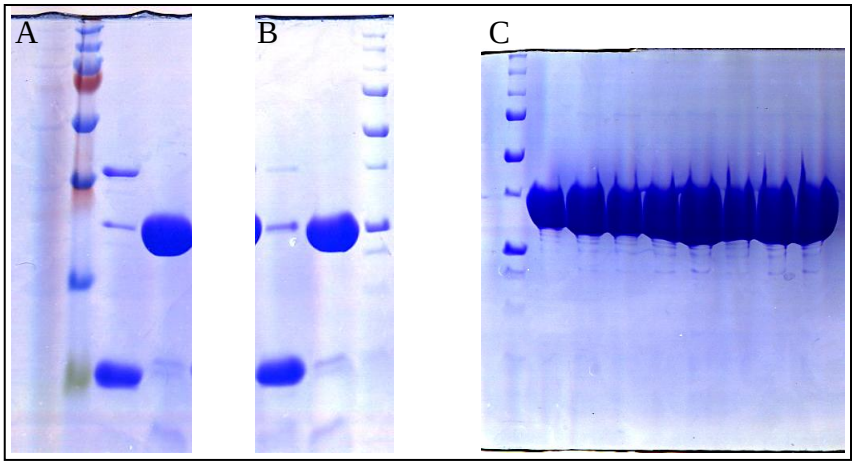


Figure 10: SDS -PAGE gel of HBHA 6_111 + CPD.

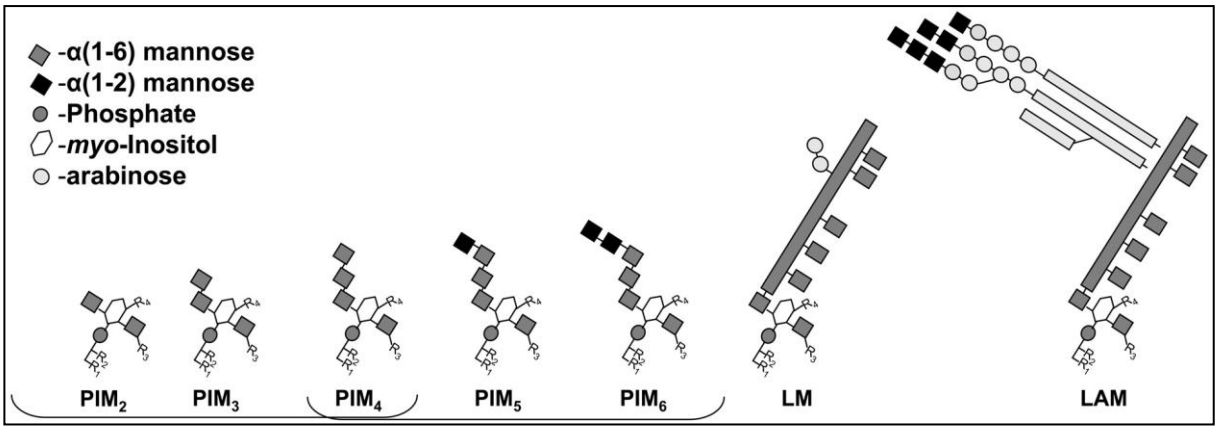


Figure 11: Structure and Biosynthesis of ManLam [37]

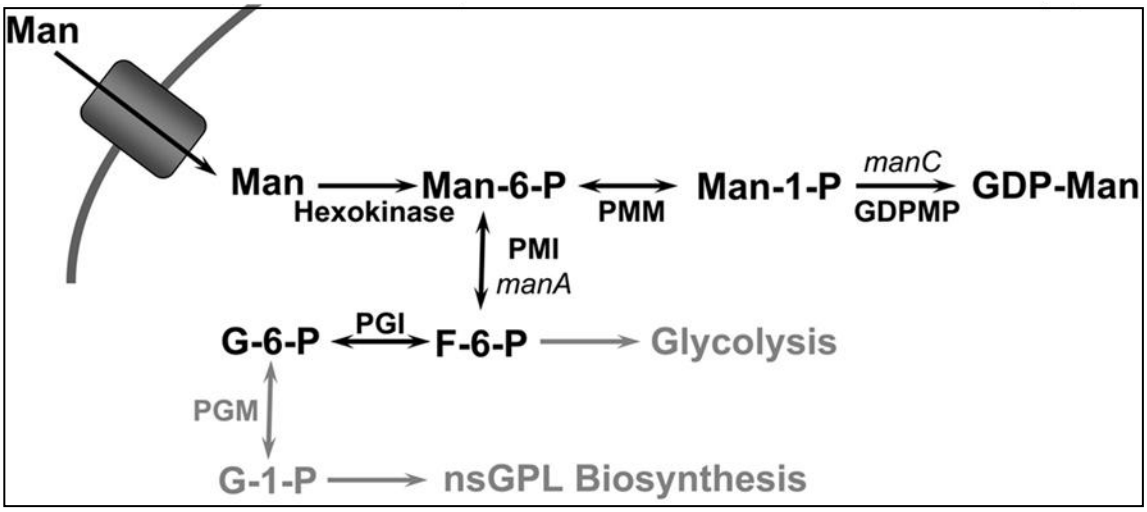


Figure 12: GDP-mannose Synthesis Pathway [37]

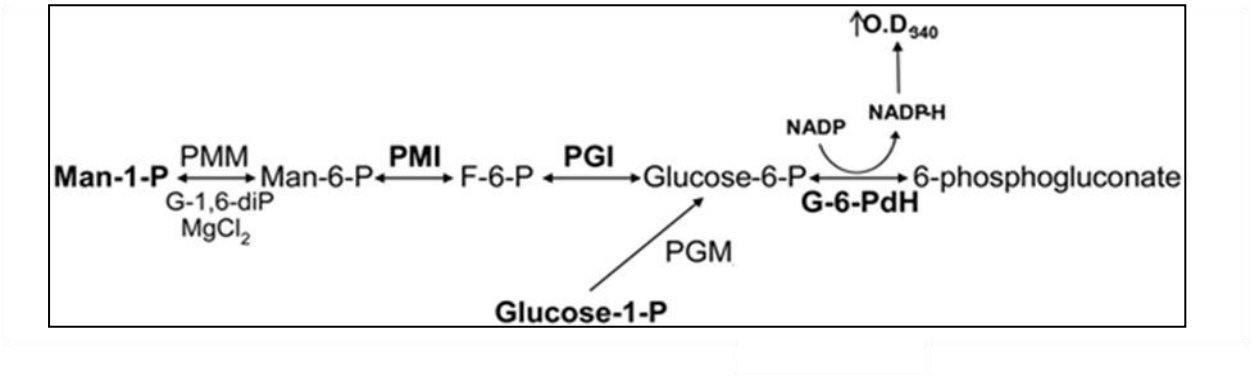


Figure 13: Phosphohexomutase Coupled Assay Schematic [37]

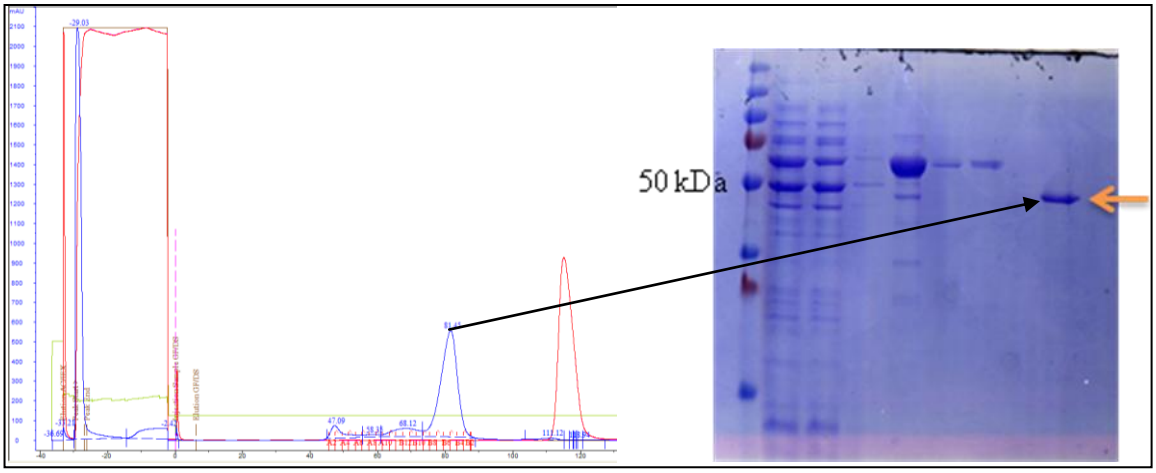


Figure 14: Size Exclusion Chromatography of Rv 3257c (Mtmanb).

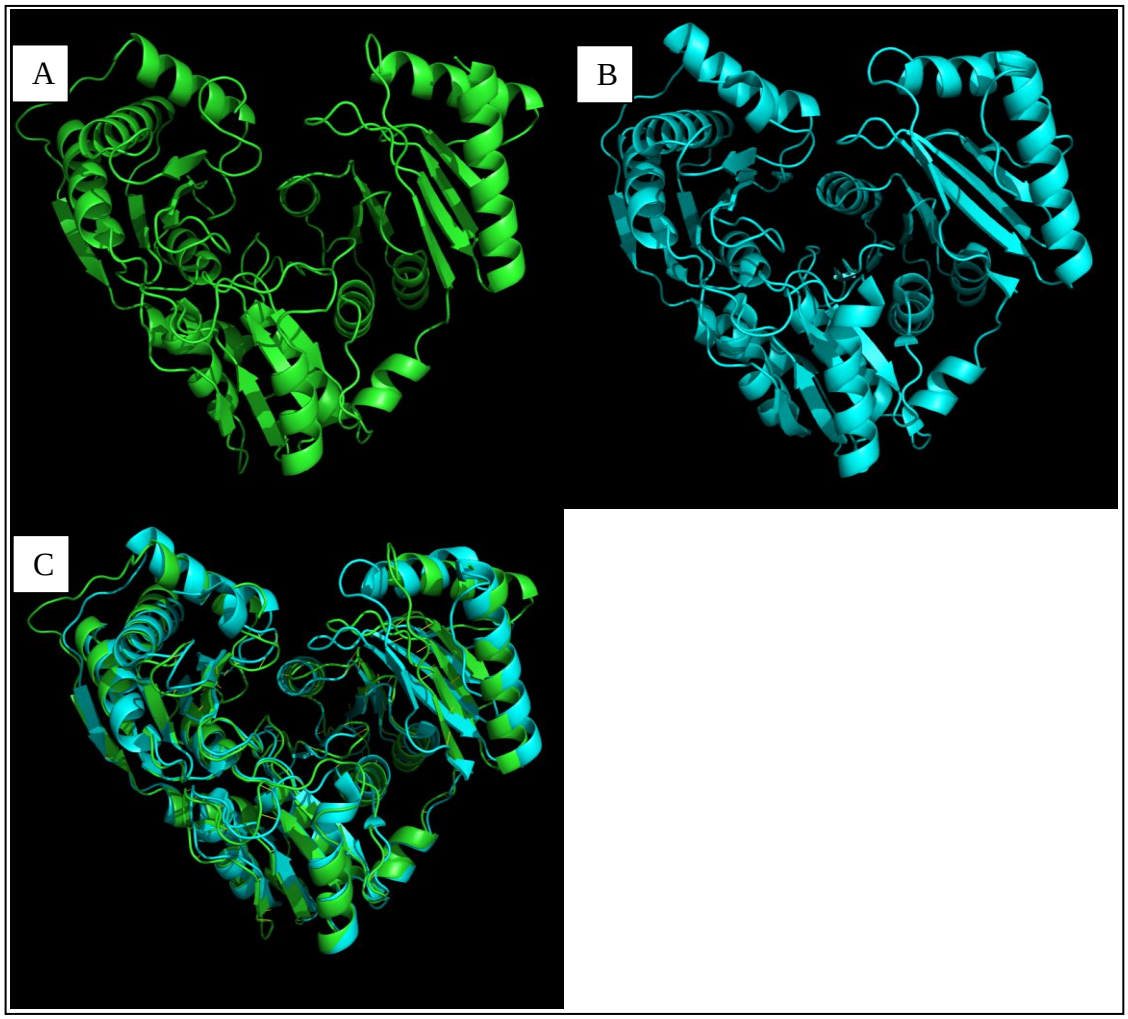


Figure 15: Ribbon Diagrams of Mt Man Model and PDB Structure 2FKF.

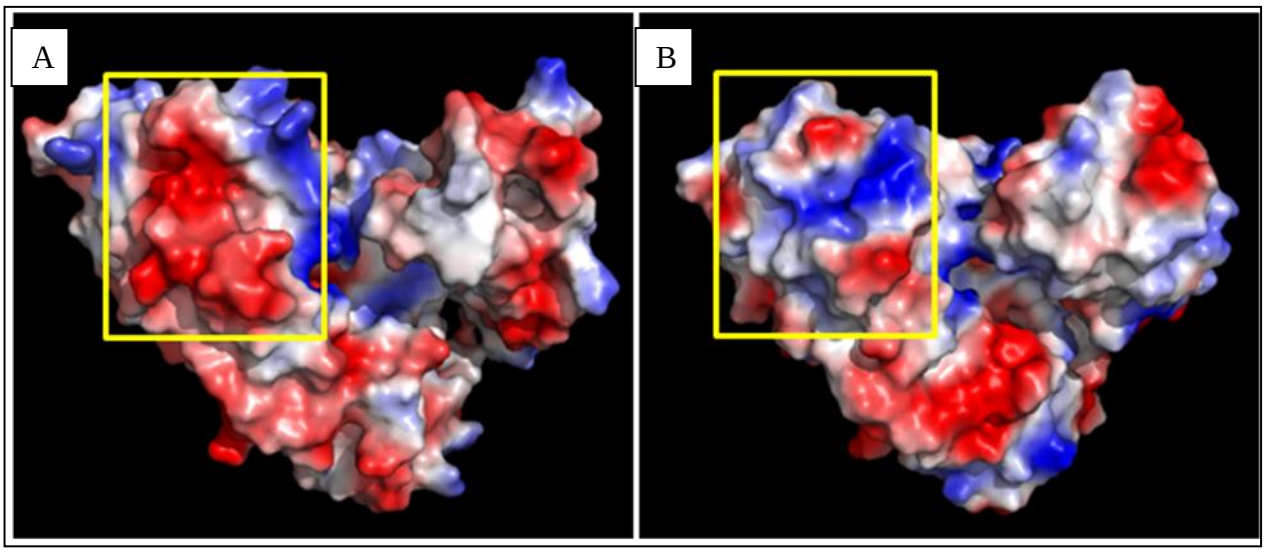


Figure 16: Charge potentials diagrams from Mtmanb model and 2FKF from the PDB

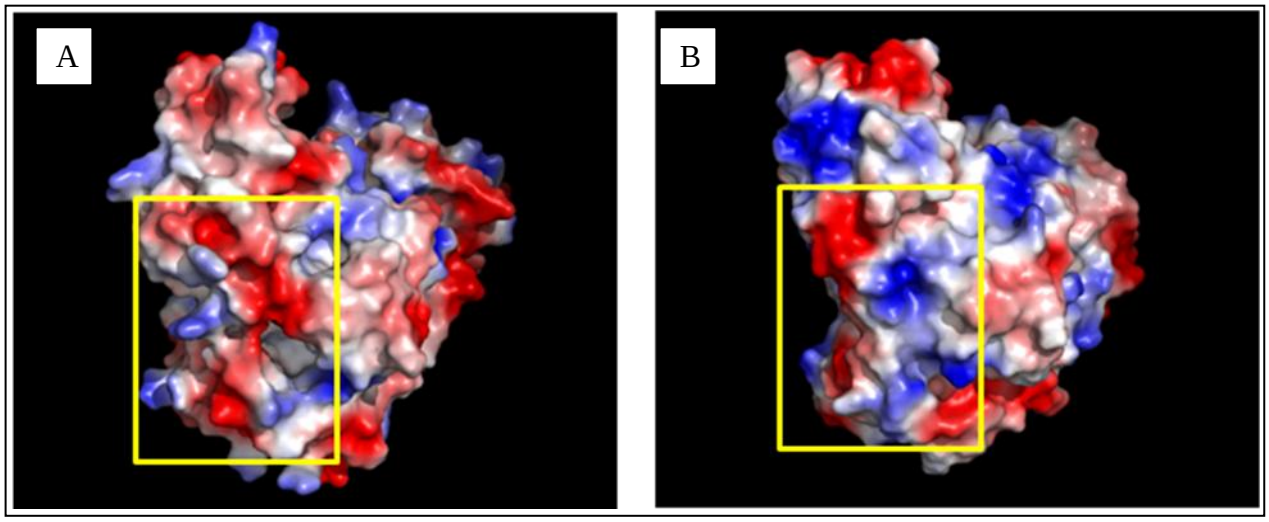


Figure 17: Charge potential diagrams from Mtmanb model and 2FKF rotated 90 degrees about the z-axis.

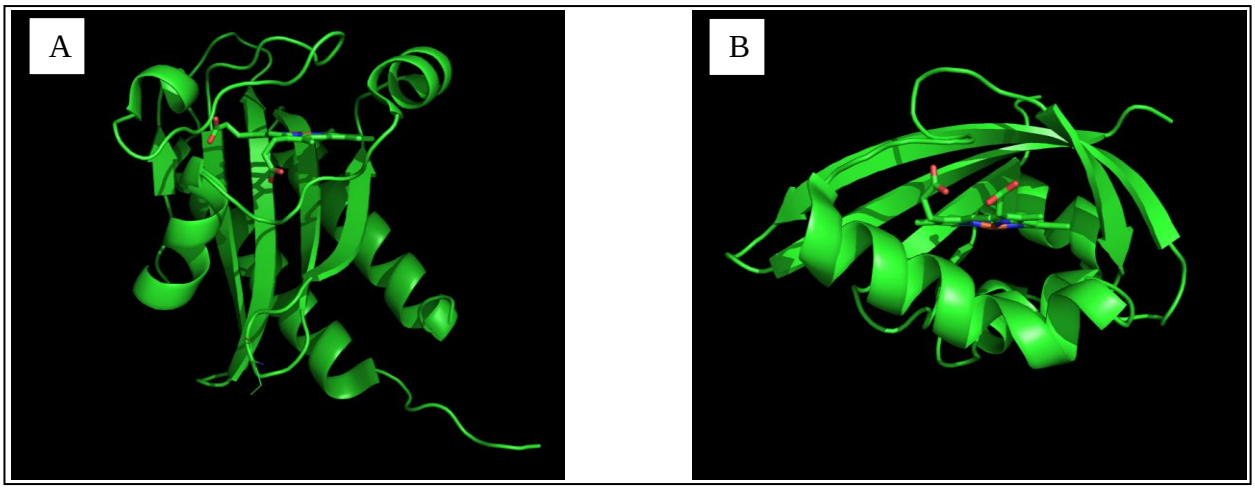


Figure 1: (A) GAF and (B) PAS domains with heme bound.

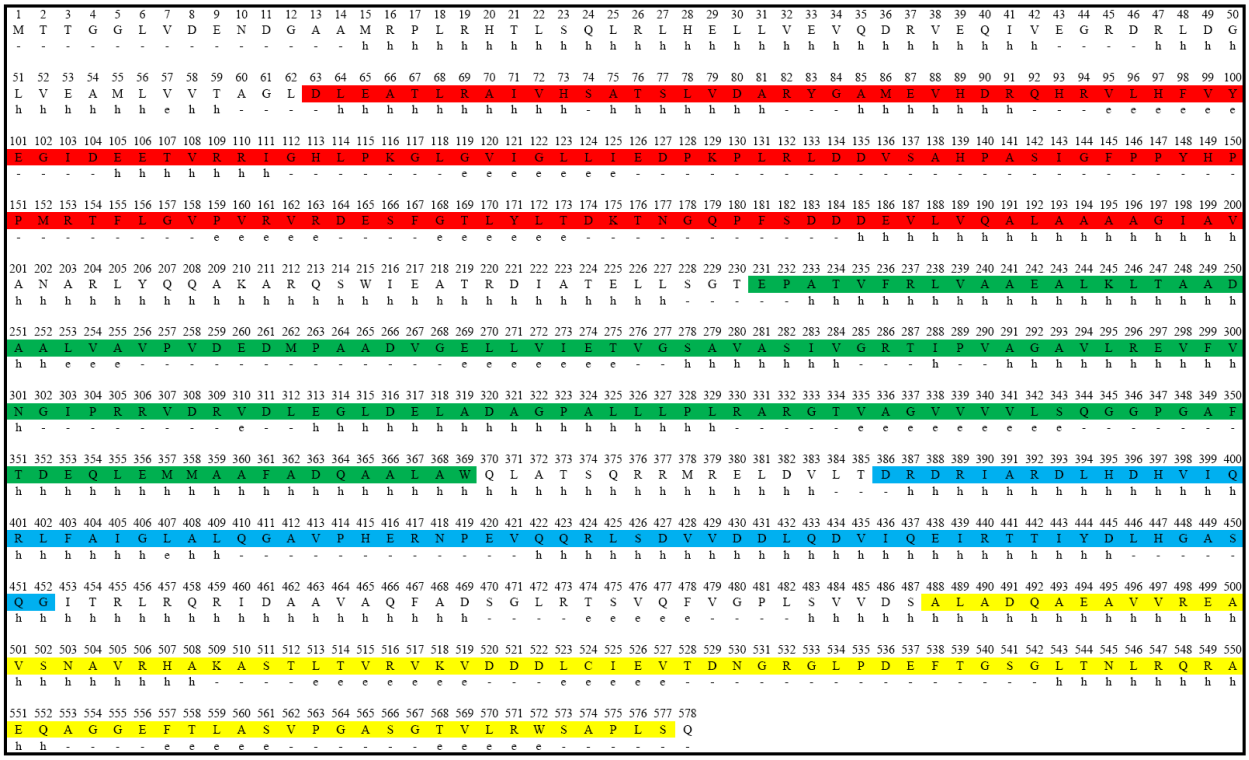


Figure 19: Secondary Structure Prediction Alignment of DevS

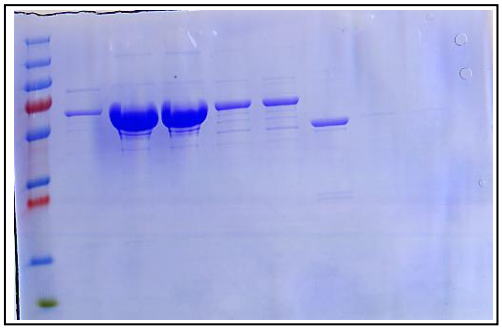


Figure 20: Gel Purification of DevS 62_578

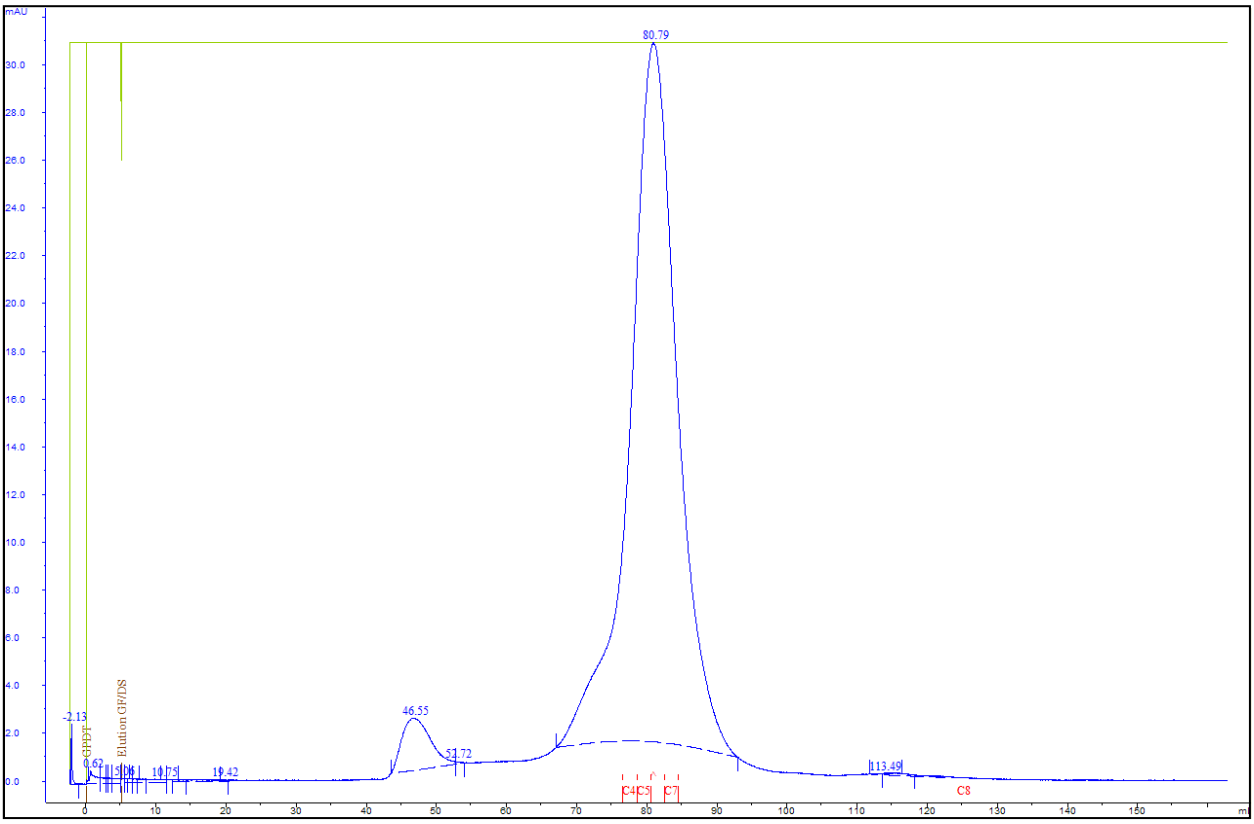


Figure 21: Size Exclusion Chromatography of DevS 62_578

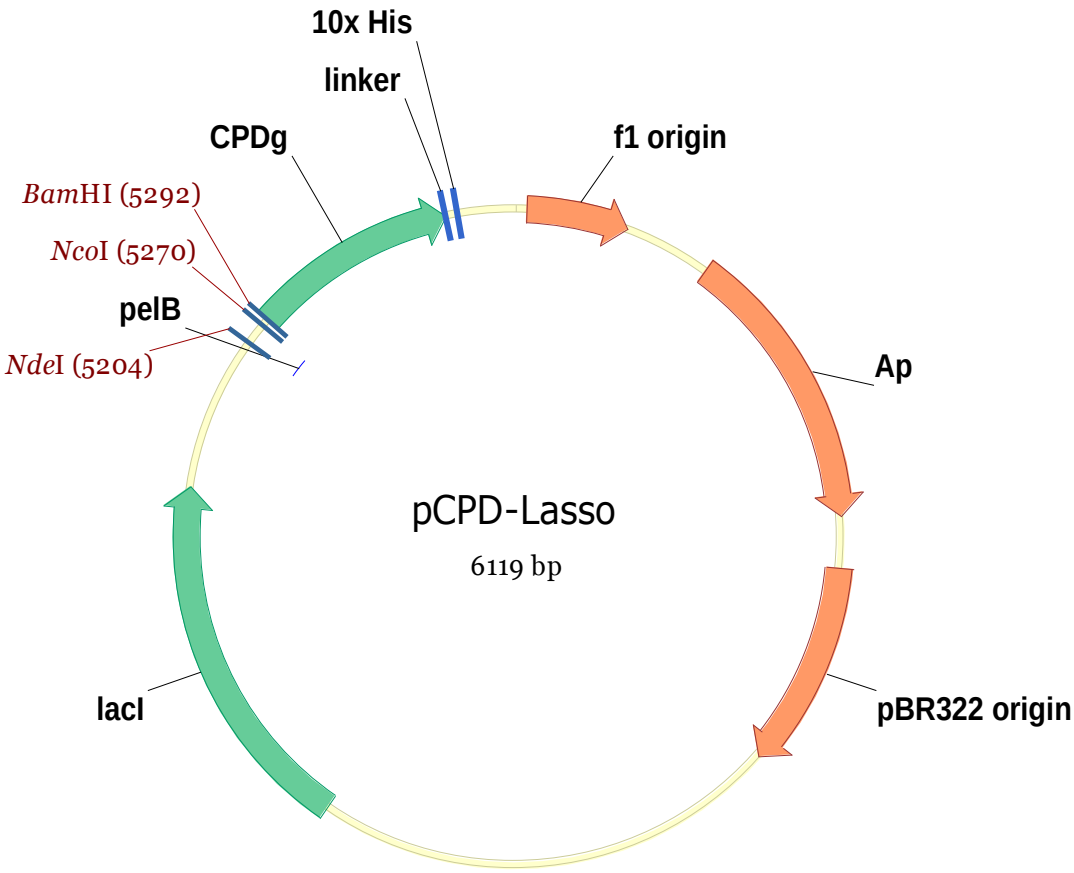


Figure 22: Vector Diagram of Assembled Vector pCPD-Lasso

Table 1: HBHA Construct List

Constructs	Residues (Start and End)	Amino Acid Number
HBHA 1_199	1 to 199	202
HBHA 6_199	6 to 199	197
HBHA 6_156	6 to 156	154
HBHA 6_111	6 to 111	109
HBHA 6_99	6 to 99	97
HBHA 6_88	6 to 88	86
HBHA 6_138	6 to 138	136
HBHA 20_199	20 to 199	183
HBHA 6_199 L14A L15A	6 to 199	197
HBHA 6_111 L14A L15A	6 to 111	109

Table 2: DICHROWEB Output (Secondary Structure Statistics for HBHA Constructs)

Constructs	Helix (%)	Sheet (%)	Turn (%)	Unordered (%)	Totals (%)
HBHA 6_111	0.62	0.05	0.1	0.22	0.99
HBHA 6_199	0.34	0.14	0.15	0.36	0.99
HBHA 6_156	0.42	0.12	0.13	0.33	1
HBHA 6_88	0.50	0.07	0.13	0.31	1.01
HBHA 20199	0.21	0.2	0.16	0.41	0.98
HBHA6_199 L14A L15A	0.16	0.28	0.16	0.41	1.01

Table 3: HBHA Construct Melting Temperatures (T_m)

Construct	Number of amino acids	Melting temperature (T_m) (^oC)
Res. 1_199	202	58.5 ± 0.20
Res. 6_199	197	60.0 ± 0.07
Res. 6_156	154	58.2 ± 0.17
Res. 6_138	136	58.0 ± 0.16
Res. 6_111	109	52.5 ± 0.16
Res. 6_99	97	40.0 ± 0.51
Res. 6_88	86	40.0 ± 1.0
Res. 6_199 L14A L15A	194	N.A.
Res. 6_111 L14A L15A	109	N.A.
Res. 20_199	183	N.A

Table 4: HBHA Construct Buoyant and Molecular Masses

Construct	Calculated Mass (Da)	Buoyant Mass (Da)	Oligomer
Res. 1_199	21675	43406	2.0
Res. 6_199	21274	49117	2.3
Res. 6_156	16973	34537	2.0
Res. 6_111	12240	26377	2.2
Res. 6_88	9867	19795	2.0
Res. 6_199 L14A L15A	21190	25350	1.2
Res. 20_199	19926	23323	1.2

Appendix A: Reference Datasets Circular Dichroism and Synchrotron Radiation Circular Dichroism Spectroscopy of Proteins[69].

Proteins included in each dataset are shown with X in the column of the data set.

Name	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	SP175
Range (nm)	178-260	178-260	185-240	190-240	178-260	185-240	190-240	175-260
AAMY		X						
ABNG	X		X	X		X	X	X†
ACY5		X	X					
ACY9		X	X					
ADH		X	X		X	X		
ADK		X		X				
ALDO		X						
APP		X						
APRT		X						
AVDN		X						
AZU	X	X	X	X		X	X	
BAMY		X						
BB2C		X						
BGAL		X						
BLAC	X	X	X	X		X	X	X†
BNJN	X	X	X	X	X	X	X	
BPTI		X		X				

CA1		X						
CANH		X	X		X	X	X	
CAL		X						
CAT		X						
CER		X						
CGA		X	X		X	X	X	
CHYT	X	X	X	X	X	X	X	X
CITS		X						
ColA		X	X		X	X		
CONA	X	X	X	X		X	X	X
CPAS		X		X	X			
CPHY		X						
CYTC	X	X	X	X	X	X	X	X†
DHQ1		X						
DHQ2		X						
DNA1		X						
ECOR	X		X	X		X	X	
ELAS	X	X	X	X	X	X	X	
FERD		X						
FLVD	X	X	X	X	X	X	X	
GCR	X		X	X		X	X	X
GDC		X						
GEC		X						
GFP		X	X		X	X		

GLOX		X						
GLUD		X						
GPB		X						
GPD	X		X	X	X	X	X	
GRS		X	X		X	X		
GSCR		X						
HAL		X						
HBN	X	X	X	X	X	X	X	X
HGDC		X						
HMRT	X	X	X	X		X	X	
HSA		X						
IFBP		X	X		X	X		
IGG		X						
INSL		X		X	X			
JAC		X						
LACF		X						
LDH	X	X	X	X	X	X	X	
LEP		X						
LLEC		X						
LYSM	X	X	X	X	X	X	X	X
MBH		X						
MGLB	X	X	X	X	X	X	X	X
MON		X						
NMRA		X						

NUCL		X		X				
OVAL		X						
OVOT		X						
OX20		X	X					
PAPN	X	X	X	X	X	X	X	X
PARV		X		X				
PELC		X						
PGK	X		X	X		X	X	X
PGLU		X						
PGM		X						
PLA2		X						
PLEC		X						
PNMT		X						
PPSN	X	X	X	X		X	X	X
PRAL	X	X	X	X	X	X	X	
PROX		X						
PYK		X						
RHD		X	X		X	X	X	
RNAS	X	X	X	X		X	X	X
RUBR		X		X				
SN06		X	X					
SN70		X	X					
STI		X						
STRP		X						

SUBA		X						
SUBB	X		X	X	X	X	X	
SUBN	X		X	X		X	X	
SUDS	X	X	X	X		X	X	X
THAU		X						
T4LS	X	X	X	X		X	X	
THML	X	X	X	X	X	X	X	
TNF	X		X	X		X	X	
TPI	X	X	X	X	X	X	X	X
TRPN		X						
UBIQ		X						

Appendix B: Protein Crystallization Screens

1. Qiagen® Screens: Classics, Classics Lite, pH Clear suite, JCSG+ Suite, JCSG 1 – 4, PACT suite, PEGs I and PEGs II suites, MPD suite, $AmS0_4$ suite

2. Molecular Dimensions Screens: Morpheus™, PGA suite, and Clear Strategy I and II.

3. Emerald BioStructures Screens: Wizards I and II

Appendix C: High-Throughput *E. coli* expression system from MARTX protein, Cysteine Protease Domain

A barrier to protein crystallization is the production of high concentrations of properly folded recombinant protein. Proteins non-native to a particular system (e.g *E. coli*) are likely not to express or to express as insoluble inclusion bodies. In this work we assembled a vector that uses an auto-cleaving molecule to help speed up purification while also increasing protein expression. Previous papers have shown how using the cysteine protease domain (CPD) of MARTX toxin as a way to purify recombinant proteins[61]. Production of many constructs to test for crystallization requires a method of screening that can cope with the volume of constructs. In building our construct, we looked at making the system more amenable to a high-throughput system.

The initial version of CPD vectors were generated for bench top columns, not for automated systems like the AKTA express. Modifications were made to accommodate for the tighter resin packing and the higher flow rates found with piston pumps systems, along with expression in *E. coli*. Codon usage is the single most important determinant in prokaryotic protein expression[62]. We analyzed the CPD DNA sequence for rare codons with respect to the *E. coli* expression system. The incidents of rare codons were not disproportionately high, but the incidences were clustered. Rare codon clustering has be shown to cause translational stop and ribosome release[63]. Use of competent cell types designed for rare codon expression, BL21 (DE3) - RIPL and BL21 (DE3) - RIP, was also seen to work to some effect, but some more recent work has shown that codon bias presented in these cells causes more insoluble expression of the recombinant protein[64]. We decided to have the CPD fragment codon optimized to express in *E. coli.* A codon optimized cDNA of CPD was ordered from GenScript™. While creating the

CPD cDNA, we also had GenScript™ remove all possible restriction sites. This facilitated insertion of cloned DNA into different expressions vectors. Along with codon optimization we expanded the hexa-his tag to a deca-his tag along with a linker. Much work has been done to determine the most effective organization for a histidine tag[65]. Labs have tried different sizes, oligo numbers, and with and without linkers[66]. A combination of a linker length around 11 with a deca-his tag was seen to be best for immobilization of the tag to a surface[67]. The deca-his tag has a greater affinity for the Nickel column allowing for higher concentrations of imidazole during load and wash steps to remove non-specifically bound material. Lastly a linker region of 10 amino acids, consisting of glycine - serine repeats, was added between the CPD and the deca-his tag to decrease the likelihood of the tag being incorporated into the fold of CPD and increasing access to the Ins6p site due to increased degrees of freedom with a flexible linker.

Design

Our CPD vector was build on the pET22b vector backbone (figure 22). The pET vectors use a T7 promoter-driven system, where control of corresponding RNA polymerase is IPTG inducible. This system allows for a level of control based upon the concentration of IPTG added. The pET22b backbone has built-in Ampicillin resistance to help with the selection process during transformation into *E. coli* for replication or protein expression. Another benefit of the pET22b vector is the presence of the pelB leader sequence. When the pelB leader sequence is coded N-terminal to the target protein the *E.coli* transports synthesis of the protein from the cytoplasm to the periplasm. Proteins that form inclusion bodies or aggregates in the cytoplasm due to the presence of disulfide bonds, fold better in the periplasm[68]. In using the pET22b there is greater flexibility in protein expression. The codon optimized CPD gene was amplified from pUC19 using PCR where the 3' primer included a sequence that would encode for a linker region

of serines and glycines (gly - gly - ser x3) along with 4 extra histidine residues for the C-terminal tag. CPD was inserted into the pET22b vector using *XhoI* and *BamHI* restriction endonucleases. A target protein can be inserted into the new vector pET22b + LASSO using two different 5' restriction sites. To remove the pelB sequence the protein is inserted using the *NdeI* site, while the *NcoI* site is used to keep the target protein within the reading frame of the pelB sequence. The *BamHI* site is used for the 3' end of the insert. When designing primers to insert the target gene it is important to look at the primary sequence of the protein. If there is a leucine near the end of a construct then the codon for leucine does not have to be included in the primer, but if no leucine is present then the leucine codon needs to be added 3' to the target protein sequence for the 3' primer. This leucine is important for CPD cleavage.

Results

Use of the CPD vector has been exclusive to the Relaxase class of proteins studied in the Redinbo. The CPD system has been shown express VirD2 (*A. tumefaciens*) and the relaxase from a vancomysin resistance plasmid (*S. aureus*). The proteins were previously expressed as inclusion bodies using the T7 promoter system in the Redinbo laboratory. Using the CPD system, the proteins remain soluble (data not shown). In all instances protein purification time is decreased from 2 - 3 days to a minimum of 7 hours in the case of VirD2.

Figure Legend:

Vector Diagram of Assembled Vector pCPD-Lasso: Vector diagram was generated using Vector NTI ™ (Invitrogen ™). Restriction sites used to insert target protein sequence are listed I red on the vector. Important features of the vector are listed in bold black. The direction of the arrows notes the direction in which the features are transcribes or replicated.

REFERENCES

[1] **N. W. Schluger, *Am J Respir Cell Mol Biol* 32 (4), 251 (2005).**

[2] **World Health Organization, 2010.**

[3] **S. Ahmad, *Clin Dev Immunol* 2011, 814943 (2010).**

[4] **P. B. Kang, A. K. Azad, J. B. Torrelles et al., *J Exp Med* 202 (7), 987 (2005); J. Nigou, C. Zelle-Rieser, M. Gilleron et al., *J Immunol* 166 (12), 7477 (2001).**

[5] **D. G. Russell, *Nat Rev Microbiol* 5 (1), 39 (2007).**

[6] **J. M. Davis and L. Ramakrishnan, *Cell* 136 (1), 37 (2009).**

[7] **A. G. Kinhikar, I. Verma, D. Chandra et al., *Mol Microbiol* 75 (1), 92 (2010).**

[8] **K. Pethe, S. Alonso, F. Biet et al., *Nature* 412 (6843), 190 (2001).**

[9] **F. D. Menozzi, R. Bischoff, E. Fort et al., *Proc Natl Acad Sci U S A* 95 (21), 12625 (1998).**

[10] **K. Pethe, V. Puech, M. Daffe et al., *Mol Microbiol* 39 (1), 89 (2001).**

[11] **C. Verbelen, V. Dupres, D. Raze et al., *J Bacteriol* 190 (23), 7614 (2008).**

[12] **J. M. Hougardy, K. Schepers, S. Place et al., *PLoS One* 2 (10), e926 (2007).**

[13] **C. Locht, J. M. Hougardy, C. Rouanet et al., *Tuberculosis (Edinb)* 86 (3-4), 303 (2006).**

[14] **G. G. Guerrero, A. S. Debrie, and C. Locht, *Vaccine* 28 (27), 4340 (2010).**

[15] **V. Dupres, F. D. Menozzi, C. Locht et al., *Nat Methods* 2 (7), 515 (2005).**

[16] **C. Esposito, P. Carullo, E. Pedone et al., *FEBS Lett* 584 (6), 1091 (2010).**

[17] **A. Lupas, *Trends Biochem Sci* 21 (10), 375 (1996).**

[18] **G. Delogu and M. J. Brennan, *J Bacteriol* 181 (24), 7464 (1999).**

[19] **K. Pethe, M. Aumercier, E. Fort et al., *J Biol Chem* 275 (19), 14273 (2000).**

[20] **W. T. Mooij, E. Mitsiki, and A. Perrakis, *Nucleic Acids Res* 37 (Web Server issue), W402 (2009).**

[21] **L. Stols, M. Gu, L. Dieckman et al., *Protein Expr Purif* 25 (1), 8 (2002).**

[22] **T. M. Cooper and R. W. Woody, *Biopolymers* 30 (7-8), 657 (1990).**

[23] **L. Whitmore and B. A. Wallace, *Nucleic Acids Res* 32 (Web Server issue), W668 (2004).**

[24] **J. M. Scholtz, H. Qian, E. J. York et al., *Biopolymers* 31 (13), 1463 (1991).**

[25] **P. Schuck, *Biophys J* 78 (3), 1606 (2000).**

[26] **P. Schuck, *Anal Biochem* 320 (1), 104 (2003).**

[27] **C. Esposito, M. V. Pethoukov, D. I. Svergun et al., *J Bacteriol* 190 (13), 4749 (2008).**

[28] **H. A. Watkins and E. N. Baker, *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64 (Pt 8), 746 (2008).**

[29] **F. D. Menozzi, J. H. Rouse, M. Alavi et al., *J Exp Med* 184 (3), 993 (1996).**

[30] **R. A. Kammerer, T. Schulthess, R. Landwehr et al., *Proc Natl Acad Sci U S A* 95 (23), 13419 (1998).**

[31] **Cummings R Varki A, Esko J, Hudson F, Hart G, and Marth J, *Essentials of Glycobiology*, 1st ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).**

[32] **R. A. Fratti, J. Chua, I. Vergne et al., *Proc Natl Acad Sci U S A* 100 (9), 5437 (2003).**

[33] **C. A. Janeway, Jr. and R. Medzhitov, *Annu Rev Immunol* 20, 197 (2002).**

[34] **A. Aderem and D. M. Underhill, *Annu Rev Immunol* 17, 593 (1999).**

[35] **Z. A. Malik, C. R. Thompson, S. Hashimi et al., *J Immunol* 170 (6), 2811 (2003).**

[36] **J. Kordulakova, M. Gilleron, K. Mikusova et al., *J Biol Chem* 277 (35), 31335 (2002).**

[37] **T. R. McCarthy, J. B. Torrelles, A. S. MacFarlane et al., *Mol Microbiol* 58 (3), 774 (2005).**

[38] **C. Regni, P. A. Tipton, and L. J. Beamer, *Structure* 10 (2), 269 (2002).**

[39] **L. E. Naught and P. A. Tipton, *Biochemistry* 44 (18), 6831 (2005).**

[40] **C. Regni, A. M. Schramm, and L. J. Beamer, *J Biol Chem* 281 (22), 15564 (2006).**

[41] **J. B. Dai, Y. Liu, W. J. Ray, Jr. et al., *J Biol Chem* 267 (9), 6322 (1992).**

[42] **L. A. Kelley and M. J. Sternberg, *Nat Protoc* 4 (3), 363 (2009).**

[43] **LLC Schrodinger, PyMOL Molecular Graphics System (2009-2010).**

[44] **H. Yamada, T. Tamada, M. Kosaka et al., *Protein Sci* 16 (7), 1389 (2007).**

[45] **I. L. Bartek, R. Rutherford, V. Gruppo et al., *Tuberculosis (Edinb)* 89 (4), 310 (2009).**

[46] **L. Nguyen and J. Pieters, *Trends Cell Biol* 15 (5), 269 (2005).**

[47] **J. M. Lee, H. Y. Cho, H. J. Cho et al., *J Bacteriol* 190 (20), 6795 (2008).**

[48] **D. M. Roberts, R. P. Liao, G. Wisedchaisri et al., *J Biol Chem* 279 (22), 23082 (2004).**

[49] **D. R. Sherman, M. Voskuil, D. Schnappinger et al., *Proc Natl Acad Sci U S A* 98 (13), 7534 (2001).**

[50] **N. Dasgupta, V. Kapur, K. K. Singh et al., *Tuber Lung Dis* 80 (3), 141 (2000).**

[51] **M. I. Voskuil, K. C. Visconti, and G. K. Schoolnik, *Tuberculosis (Edinb)* 84 (3-4), 218 (2004).**

[52] **Mayuri, G. Bagchi, T. K. Das et al., *FEMS Microbiol Lett* 211 (2), 231 (2002).**

[53] **Y. S. Ho, L. M. Burden, and J. H. Hurley, *Embo J* 19 (20), 5288 (2000).**

[54] **S. Sardiwal, S. L. Kendall, F. Movahedzadeh et al., *J Mol Biol* 353 (5), 929 (2005).**

[55] **H. Y. Cho, H. J. Cho, Y. M. Kim et al., *J Biol Chem* 284 (19), 13057 (2009).**

[56] **E. T. Yukl, A. Ioanoviciu, M. M. Nakano et al., *Biochemistry* 47 (47), 12532 (2008).**

[57] **H. Y. Cho, H. J. Cho, Y. M. Kim et al., *Acta Crystallogr Sect F Struct Biol Cryst Commun* 64 (Pt 4), 274 (2008).**

[58] **E. T. Yukl, A. Ioanoviciu, P. R. de Montellano et al., *Biochemistry* 46 (34), 9728 (2007).**

[59] **J. D. Bendtsen, H. Nielsen, G. von Heijne et al., *J Mol Biol* 340 (4), 783 (2004).**

[60] **M. Ferrer, T. N. Chernikova, M. M. Yakimov et al., *Nat Biotechnol* 21 (11), 1266 (2003).**

[61] **A. Shen, P. J. Lupardus, M. Morell et al., *PLoS One* 4 (12), e8119 (2009).**

[62] **G. Lithwick and H. Margalit, *Genome Res* 13 (12), 2665 (2003).**

[63] **T. F. th Clarke and P. L. Clark, *PLoS One* 3 (10), e3412 (2008).**

[64] **G. L. Rosano and E. A. Ceccarelli, *Microb Cell Fact* 8, 41 (2009).**

[65] **M. He and M. J. Taussig, *Nucleic Acids Res* 29 (15), E73 (2001).**

[66] **S. Knecht, D. Ricklin, A. N. Eberle et al., *J Mol Recognit* 22 (4), 270 (2009).**

[67] **M. Fischer, A. P. Leech, and R. E. Hubbard, *Anal Chem* (2011).**

[68] **M. Schlapschy and A. Skerra, *Methods Mol Biol* 705, 211 (2011).**

[69] **L. Whitmore, B. Woollett, A. J. Miles et al., *Structure* 18 (10), 1267 (2010).**