Determining the Role of the *Mycobacterium tuberculosis* Serine/Threonine Protein Kinase, PknH, in Cell Signaling

by

Alexandra Cavazos

A dissertation submitted in partial satisfaction of the requirements of the degree of

Doctor of Philosophy in Molecular and Cell Biology

in the Graduate Division of the University of California, Berkeley

Committee in Charge:
Prof. Tom Alber, Chair
Prof. James Berger
Prof. Steven Martin
Prof. Fenyong Liu

Spring 2012

ABSTRACT

Determining the Role of the *Mycobacterium tuberculosis* Serine/Threonine Protein Kinase, PknH, in Cell Signaling

by
Alexandra Cavazos
Doctor of Philosophy in Molecular and Cell Biology
University of California, Berkeley
Professor Tom Alber, Chair

The survival of the pathogenic bacterium, *Mycobacterium tuberculosis* (*Mtb*), the causative agent of tuberculosis, depends on its ability to sense and respond to a hostile and constantly changing environment within the host. Cell-wall components--such as peptidoglycan, arabinogalactan, mycolic acids and lipoarabinomannan--perform critical structural and biological functions that allow *Mtb* to survive and persist inside of human cells. Despite the critical roles of the cell wall in the pathogenicity of *Mtb*, surprisingly little is known about how this complex structure is built and remodeled in response to environmental cues. Recent evidence, however, implicates membrane-bound serine/threonine protein kinases (STPKs) in coordinating cell-wall synthesis with growth and division. These eukaryotic-like STPKs play critical roles in *Mtb* signal transduction by sensing extracellular stimuli and catalyzing the reversible phosphorylation of cytoplasmic target proteins.

To further understand the mechanisms that regulate STPKs and cell wall architecture in *Mtb*, I biochemically and structurally characterized the extracellular sensor domain of one STPK receptor kinase, PknH. Initial analysis of recombinantly produced PknH sensor domain by size-exclusion chromatography and small-angle x-ray scattering indicated the protein was soluble but completely unfolded. On-column oxidative refolding produced a properly folded protein that crystallized as a monomer with a novel protein fold made up of six alpha helices, seven beta strands and two disulfide bonds. The PknH sensor domain has a large conserved cleft with a mixed polar and hydrophobic surface. These results indicated PknH binds a small molecule ligand which may affect its quarternary structure and/or its localization.

In order to determine what ligand binds to the PknH sensor domain, I conducted native gel binding assays using purified *Mtb* arabinogalactan, lipoarabinomannan, and lipomannan. These assays indicated that *in vitro*, the PknH sensor domain does not bind to any of these purified cell wall components. Affinity chromatography with His-tagged PknH incubated with *Mtb* H37Rv cell lysate followed by intact mass, LC-MS, and ESI-MS analysis indicated that PknH does not bind glycolipids such as phosphoinositol or Ac_1PIM_2. Although the ligand for PknH was not identified, these studies set the stage for future researchers to identify the ligand and discover the function of PknH.

Dedication

This doctoral dissertation is dedicated to the memory of my parents,
Maria Montez Cavazos and Alonzo Cavazos.

Acknowledgements

I want to thank Daniil M. Prigozhin, Ursula Schultze-Gamen, Daniela Mavrici, Terry Lang, Alexander Scouras, Ho-Jun Lee, Seemay Chou, Heather Upton, John Huizar, Kristi Pullen, Andrew Greenstein, James Fraser, Carl Mieczkowski, Christina Baer, Christine Gee, and Noelle Lombana of the Alber lab, and Prof. Kimmen Sjolander and the Berkeley Phylogenomics Group, for their advice and helpful discussions about my research. I also thank my thesis committee members Tom Alber, James Berger, G. Steven Martin, and Fenyong Liu for their useful feedback. I am indebted to Prof. Alber for giving me the opportunity to join his lab and learn about the world of molecular biology under his guidance.

I sincerely appreciate the mentoring I received at San Francisco State University from Dr. Frank Bayliss, Dr. Zheng-Hui He, Dr. Ray Esquerra, Dr. Carmen Domingo, and Dr. Leticia Márquez-Magaña. These professors generously shared their experiences and acquired wisdom with me, and I am very thankful for their help.

A heartfelt thank you goes to my brother, Alonzo Cavazos, Jr., and his wife, Rosalie Cavazos, for their love and assistance throughout my academic career. My family members Ruby Cavazos, Perla Cavazos, Paul Saucido, Arturo Cavazos, Dan and Helen Cavazos, Sara and Morrie Verner, and Tara Tanwongprasert also deserve special mention for giving me their unwavering support. Last but not least, I also thank Bruno Martinez, Robert Thompson, Antonio Garcia, and Jane Wong for their friendship and encouragement. I could not have made it this far without the aid of all these family members and friends; you all are my giants, on whose shoulders I gratefully stand.

TABLE OF CONTENTS

CHAPTER 1 – INTRODUCTION

I. Features of Serine/Threonine Protein Kinases (STPKs)

Protein phosphorylation is a post-translational protein modification that transduces environmental signals into appropriate cellular responses (Stock, Ninfa et al. 1989). Protein kinases catalyze protein phosphorylation and represent the largest and most diverse gene family in eukaryotes (Manning, Whyte et al. 2002). These kinases are divided into two major families: (1) the histidine kinase superfamily, which contains the two-component sensor kinases that autophosphorylate on a conserved histidine residue (Grebe and Stock 1999), and (2) the serine, threonine and tyrosine protein kinases that make up the STPK superfamily and phosphorylate substrates on serine, threonine and tyrosine residues, respectively (Hanks and Hunter 1995).

In eukaryotes, STPKs are traditionally grouped together based on sequence homology of the kinase domains, which have 12 characteristic subdomains that form a two-lobed catalytic core (Hanks and Hunter 1995) (Figure 1.1a). The phosphate donor, ATP, binds between the lobes of the catalytic domain, while the larger lobe binds the substrate protein and promotes the transfer of the phosphate group (Pereira, Goss et al. 2011). The active site sits in a deep pocket formed between the two lobes (Hanks and Hunter 1995; Kornev and Taylor 2010) and is defined by the presence of a P-loop, a catalytic loop, a magnesium-binding loop, and an activation segment (Figure 1.1b).

For many years, researchers assumed prokaryotes possessed only two-component histidine kinases (Hoch 2000), but biochemical studies and the advent of whole-genome sequencing of multiple bacterial species allowed for the identification of several eukaryotic-like serine/threonine protein kinases (STPKs) in prokaryotes in the 1990's (Hanks and Hunter 1995).

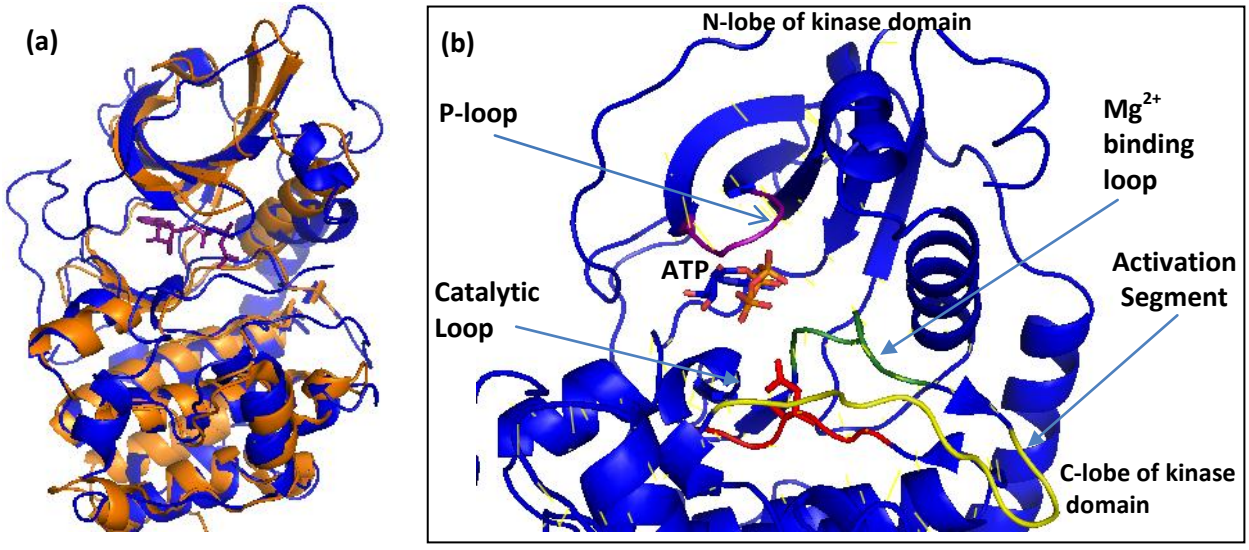


Figure 1.1 (a) Superposition of ATP-bound PKA (blue ribbon with ATP molecule in purple sticks) with the catalytic domain of *Mycobacterium tuberculosis* PknB (orange ribbon) demonstrates their structural similarities. **(b)** A close up view of the PKA kinase domain illustrates characteristic active-site elements for an STPK. The Mg^{2+}-ATP sits between the upper N-terminal lobe and the lower C-terminal lobe. Sitting above the ATP molecule is the P-loop (magenta), which forms essential contacts with the ATP phosphate groups. The catalytic loop (red) sits below the ATP molecule; the key catalytic Asp residue is shown in red sticks. The Mg^{2+}-binding loop is colored green, while the P+1 loop and the activation loop that make up the activation segment are colored yellow. This figure is adapted from Figure 1 in Pereira, Goss et al. 2011.

More recently, Global Ocean Sampling data demonstrated that eukaryotic-like STPKs outnumber histidine kinases (Kannan, Taylor et al. 2007), indicating that these signaling molecules are at least as important in prokaryotic cellular regulation as histidine kinases.

II. The architecture and functions of *Mtb* STPKs with extracellular sensor domains

Mycobacterium tuberculosis (*Mtb*), the causative agent of tuberculosis, has 11 histidine-kinase/response-regulator pairs, along with 11 STPKs, several Ser/Thr phosphatases, and 1 Tyr phosphatase (Cole, Brosch et al. 1998; Av-Gay and Everett 2000; Greenstein, Cavazos et al. 2009 (Appendix A); Ortiz-Lombardia and Pompeo 2003). Nine of the 11 *Mtb* STPKs have a single predicted transmembrane helix (Figure 1.2), while the other two, PknG and PknK, have no transmembrane helices.

The extracellular sensor domain of PknB has four PASTA domains arranged in a linear fashion (Barthe, Mukamolova et al. 2010), while the PknD sensor has six NHL domains that form a six-bladed beta-propeller (Good, Greenstein et al. 2004). The structure for the sensor domain of PknE has not been solved, but it is predicted to adopt a DsbG-like, disulfide isomerase fold (Gay, Ng et al. 2006). Until this study was completed, the sensor domain of PknH was uncharacterized, but it had been predicted to be folded due to its weak sequence identity to the *Mtb* lipoprotein LppH.

A. PknA, PknB and PknL form a clade that regulates peptidoglycan synthesis, cell shape and cell division

Phylogenetic analysis of the *Mtb* STPKs indicates PknA, PknB and PknL are conserved across mycobacterial species and form a single evolutionary clade (Narayan, Sachdeva et al.

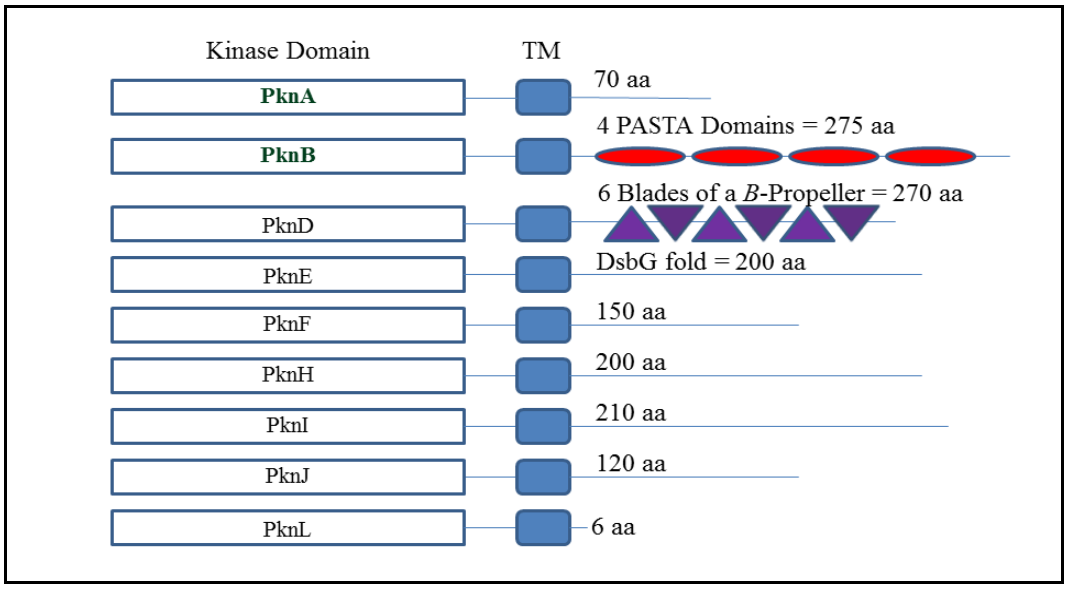


Figure 1.2 Domain architecture of the nine transmembrane STPKs in *Mycobacterium tuberculosis*. The two essential kinases, PknA and PknB, are labeled in green. The intracellular kinase domain of each STPK is boxed in white, while the transmembrane helices are boxed in blue. Two of the TM kinases, PknB and PknD, have extracellular domains that are known to be folded, while the PknE sensor domain is predicted to be folded due to its homology to the *E. coli* disulfide isomerase DsbG.

2007). Phenotypic analysis of cells with altered expression levels of PknA and PknB also indicates these essential STPKs play a role in cell growth and division. Overexpression of PknA and PknB in *Mycobacterium smegmatis* and *Mycobacterium bovis* leads to cells with defective shapes, while partial knockdown results in elongated cells (Kang, Abbott et al. 2005). The presence of four PASTA domains in PknB indicates this STPK likely sense peptidoglycan fragments and regulates cell-wall remodeling and cell division. The PknB homolog in *Bacillus subtilis*, PrkC, binds meso-diaminopimelic acid (mDAP)-containing peptidoglycan fragments, and these fragments act as powerful germinants for *B. subtilis* spores (Shah, Laaberki et al. 2008). In *Mtb*, the extracellular PASTA domains of PknB bind synthetic muropeptides in a way that depends on both the presence of specific amino acids on the stem peptide and on the sugar moiety, N-acetylmuramic acid, linked to the peptide (Mir, Asong et al. 2011). PknB localizes strongly to the mid-cell and to the cell poles, and its extracellular domain is required for its proper localization (Mir, Asong et al. 2011). The synthetic muropeptide that bound PknB with the highest affinity had a moderate effect on resuscitating dormant cells, although this effect was small compared to the effect of conditioned medium (Mir, Asong et al. 2011). In light of these findings, researchers suggest that PknB is targeted to the sites of peptidoglycan turnover to regulate cell growth and cell division.

A further indication that PknA and PknB play important roles in regulating cell shape and division in *Mtb* arises from an analysis of their identified substrates. PknA and PknB phosphorylate Wag31 (Rv2145c), an essential mycobacterial homolog of DivIVA. In *Streptomyces*, *Corynebacterium*, and *Mycobacterium*, DivIVA controls polar cell-wall synthesis (Nguyen, Scherr et al. 2007; Kang, Nyayapathy et al. 2008). In *Corynebacterium glutamicum*, DivIVA is required for cell elongation and is recruited to the septum only after peptidoglycan synthesis related to cell division has already begun (Letek, Ordóñez et al. 2008). Another indication that Wag31 plays a role in cell shape maintenance lies in the fact that changes in the PknA phosphorylation site on Wag31 altered cell shapes (Kang, Abbott et al. 2005). Changes in the activity or localization of Wag31 via phosphorylation may explain the phenotypes observed in mycobacterial cells with altered PknA or PknB expression levels (Molle and Kremer, 2010).

PknA has also been shown to phosphorylate MurD, a cytoplasmic enzyme that catalyzes the addition of D-glutamate to UDP-N-acteylmuramoyl-L-alanine precursor during peptidoglycan synthesis (Thakur and Chakraborti 2006), and FstZ, a central protein in bacterial septum formation. PknA phosphorylation of FtsZ reduces its GTP-dependent polymerization and thus reduces septum formation (Thakur and Chakraborti 2006). More recently, PknB has been shown to phosphorylate MviN, the flippase that transports building blocks of peptidoglycan synthesis across the plasma membrane (Gee, Papavinasasundaram et al. 2012).

While the data indicating PknA and PknB regulate cell shape and division is considerable, the evidence supporting a similar role for PknL is less developed. *pknL*, like *pknA* and *pknB*, is conserved across all mycobacteria (Narayan, Sachdeva et al. 2007). *pknL* is located close to the division cell wall (*dcw*) cluster of genes that control cell wall synthesis and cell division (Narayan, Sachdeva et al. 2007), and it also neighbors Rv2175c, which encodes a DNA-binding transcriptional regulator that is also an *in vitro* substrate for PknL. PknL phosphorylates Rv2175c and inhibits its binding to DNA (Cohen-Gonsaud, Barthe et al. 2009), although the specific genes regulated by Rv2175c have yet to be determined. For now, the evidence that PknL regulates cell wall synthesis or cell division remains speculative.

B. PknD, PknE and PknH form a clade of undetermined function

PknD, PknE, and PknH form another evolutionary clade of STPKs. As indicated in Figure 1.3, the PknD sensor domain forms a highly symmetrical six-bladed beta-propeller (Good, Greenstein et al. 2004). Unfortunately no clues about PknD function can be gleaned

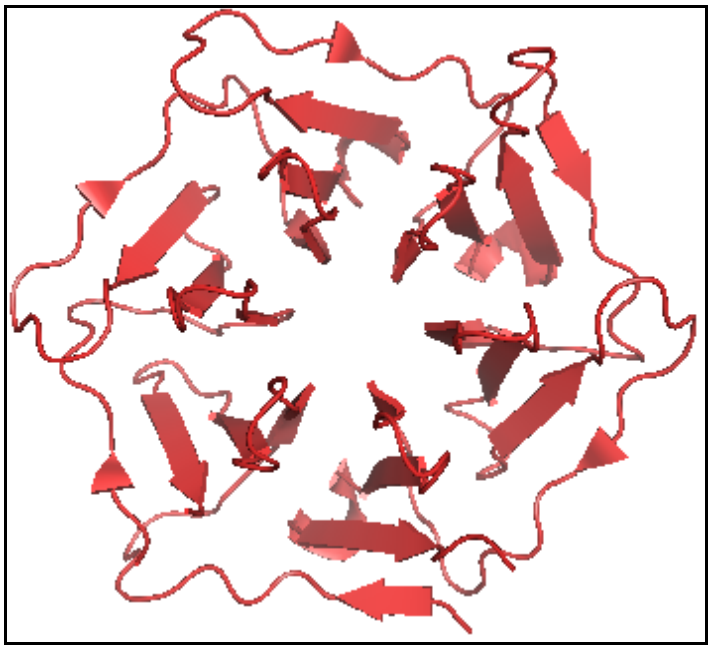


Figure 1.3 Ribbon representation of the extracellular sensor domain of the *Mtb* STPK PknD.

from the structure of its sensor domain, because beta-propellers bind a large and diverse group of ligands and perform many different catalytic functions (Pons, Gomez et al. 2003; Fulop, Jones et al. 1999). Phenotypic clues about the function of this STPK are also unavailable, as altering the levels of PknD expression by gene deletion produces no growth phenotype in culture or in mice (Parish, Smith et al. 2003). Our lab has discovered that PknD phosphorylates the anti-anti-sigma factor Rv0516c on Thr2, which blocks its binding to the anti-sigma factor Rv2638 (Greenstein, MacGurn et al. 2007), but further discovery of PknD substrates and ligands is necessary for a definitive assignment of its function.

The STPK PknE has a unique sensor domain not found together with any kinase outside of pathogenic mycobacteria (Gay, 2007). It is approximately 200 amino acids in length and has homology to the protein disulfide isomerase DsbG (Molle, Girard-Blanc et al. 2003). The PknE promoter responds to nitric oxide stress and deletion of PknE in *Mtb* results in increased apoptosis in the human macrophage *in vitro* model of infection (Jayakumar, Jacobs et al. 2008). Recent studies using DNA microarrays with wild-type and PknE knockout strains of *Mtb* indicate PknE responds to nitric oxide stress and contributes to defective apoptosis by altering the expression of Toll-like receptors that in turn suppress an intrinsic mode of apoptosis (Kumar and Narayanan 2012). This impaired apoptosis elevates pro-inflammatory responses and modulates co-stimulatory molecule expression; additionally, researchers found that increased levels or arginase1 and phosphorylation of the survival kinase Akt also contributed to suppression of apoptotic signaling (Kumar and Narayanan 2012).

C. PknH regulates key arabinosyltransferases responsible for arabinogalactan and lipoarabinomannan synthesis

PknH is a non-essential but nonetheless critical regulatory protein in *Mtb*. The *Mtb* PknH knockout strain displays a significantly higher bacillary load in the mouse infection model (Papavinasasundaram, Chan et al. 2005). The PknH deletion strain tested by Papavinasasundaram and coworkers was made via allelic exchange and then intravenously injected into 6-8 week old BALB/C mice. At the onset of adaptive immunity, the mice had 20-fold more bacteria in their lungs and spleen than control mice. As a control, a complementation assay was successfully performed with PknH expressed from its native promoter (Papavinasasundaram, Chan et al. 2005). In liquid culture, a PknH knockout strain was more resistant to acidified –nitrite stress, indicating PknH may regulate growth in response to nitric oxide stress *in vivo* (Papavinasasundaram, Chan et al. 2005). Additionally, other researchers have shown via RT-PCR on RNA isolated from *Mtb* cultures that PknH transcript levels decrease after exposure to low pH (4.5) or heat shock (42 degrees) (Sharma, Chandra et al. 2004).

The idea that PknH plays a key role in cell-wall synthesis is foremost supported by its phosphorylation of EmbR *in vitro* (Molle, Kremer et al. 2003). EmbR belongs to the *Streptomyces coelicolor* antibiotic regulatory protein (SARP) family (Wietzorrek and Bibb 1997) that regulates genes involved in secondary metabolite synthesis. The three-dimensional structure of EmbR shows it possesses an FHA domain, a small module capable of mediating protein-protein interactions through recognition of phosphorylated threonine residues, along with an N-terminal winged helix-turn-helix DNA binding domain, followed by a central bacterial transcription activation domain, suggesting it acts as a transcriptional regulator in *Mtb* (Alderwick, Molle et al. 2006). Phosphorylation of EmbR by PknH enhances its binding to the promoter region of the *EmbCAB* genes in *Mtb* (Sharma, Gupta et al. 2006b), and deletion of PknH in *Mtb* results in decreased transcription of *embB* and *embC* genes in cultures treated with sublethal concentrations of ethambutol (Papavinasasundaram, Chan et al. 2005).

The EmbCAB gene products are arabinosyltransferases with segregated biological functions; EmbA and EmbB make the crucial terminal hexaarabinoside motif in the cell-wall polysaccharide arabinogalactan (Escuyer, Lety et al. 2001), while EmbC is involved in the biosynthesis of the arabinan portion of lipoarabinomannan (LAM) (Zhang, Torrelles et al. 2003). LAM is an important determinant of virulence because it suppresses the host immune system, whereas its precursor lipomannan represents a potent pro-inflammatory factor (Briken, Porcelli et al. 2004). Therefore, by increasing the LAM/LM balance, the PknH/EmbR pair is likely to play an important role in immunopathogenesis.

EmbR is also phosphorylated by PknA and PknB and dephosphorylated by PstP, indicating that several pathways might be involved in generating a global response by integrating diverse signals (Sharma, Gupta et al., 2006a). Moreover, PknH is upregulated *in vivo* inside host macrophages (Sharma, Chandra et al. 2004), which suggests a biologically relevant signaling mechanism exists to modulate arabinan synthesis.

D. PknH and other STPKs phosphorylate enzymes involved in synthesizing mycolic acid precursors

Further evidence in support of a role for PknH in coordinating cell-wall remodeling with cell division stems from its phosphorylation of key enzymes that synthesize mycolic acid

precursors. Mycolic acids are a homologous series of C_{60}-C_{90} α-alkyl-β-hydroxy fatty acids produced by all mycobacteria (Kremer, Douglas et al. 2000) that are typically over 60 carbons in length and are important for the growth, survival and pathogenicity of *Mtb*.

Mycolic acid biosynthesis involves two types of fatty acid-synthesizing enzymes, the type I and type II fatty acid synthases FAS-I and FAS-II, respectively (Marrakchi, Bardou et al. 2008). The eukaryotic-like FAS-I system catalyzes the *de novo* synthesis of fatty acids from acetyl-CoA (Marrakchi, Bardou et al. 2008). In contrast, FAS-II is similar to systems found in bacteria, apicomplexal parasites and plants, and is composed of several enzymes that act successively and repetitively to elongate the growing acyl chain (Bhatt, Molle et al. 2007). Acyl-primers are repeatedly activated via a thioester linkage to the prosthetic group of coenzyme A (CoA) for FAS-I, or of an acyl carrier protein for FAS-II (Bhatt, Molle et al. 2007).

The biosynthesis occurs in five separate stages (Takayama, Wang et al. 2005). It begins with the synthesis of long-chain acyl-CoA with a bimodal distribution, C_{16}-C_{18} and C_{24}-C_{26}, which is a unique feature of mycobacterial FasI (Bloch and Vance, 1977; Fernandez and Kolattukudy, 1996). The saturated C_{26} chains become the alpha-alkyl branch of the mycolic acids. Next is the synthesis of the meromycolic acid by FAS-II, followed by the addition of functional groups to the meromycolate chain by several cyclopropane synthases (Takayama, Wang et al. 2005). This leads to the final stages in the reaction, which are the condensation reaction between the alpha-branch and the meromycolate chain, a final reduction to generate a mycolic acid, and the transfer of mycolic acids to arabinogalactan and other acceptors such as trehalose (Bhatt, Molle et al. 2007).

The reaction that leads to carbon-carbon bond formation during acyl-group elongation is catalyzed by condensing enzymes of three different types. The initiation condensing enzyme *Mt*FabH links the FAS-I and FAS-II systems and uses acyl-CoA primers (Choi, Kremer et al. 2000), whereas the elongating condensing enzymes KasA and KasB exclusively use acyl-ACP thioesters (Schaeffer, Agnihotri et al. 2001; Kremer, Dover et al. 2002). The termination condensing enzyme Pks13 uses both a meromycolyl-AMP and an acyl-CoA as a substrate (Portevin, De Sousa-D'Auria et al. 2004). Mycolic acids confer important properties, including resistance to chemical injury and dehydration, low permeability to hydrophobic antibiotics, virulence (Dubnau, Chan et al. 2000; Glickman, Cox et al. 2000; Glickman and Jacobs, 2001), and persistence within the host (Daffe and Draper 1998; Yuan, Zhu et al. 1998; Bhatt, Fujiwara et al. 2007).

Molle and coworkers propose a regulatory model of mycolic acid synthesis based on STPK phosphorylation of the beta-ketoacyl-ACP synthases KasA and KasB (Molle, Kremer et al. 2010). Both enzymes are phosphorylated at high rates *in vivo* in *M. bovis* BGC and are dephosphorylated by PstP (Molle, Brown et al. 2006). Phosphorylation decreases the activity of KasA but increases the activity of KasB in the presence of malonyl-ACP or C_{16}-ACP (Molle, Brown et al. 2006). The disparate effect of phosphorylation of these two enzymes, which share the same enzymatic activity but have different substrate specificities, might allow for dampening of KasA in order to produce immature mycolates while increasing KasB activity might ensure that only full-length mycolates required for virulence and survival are produced (Bhatt, Fujiwara et al. 2007; Bhatt, Molle et al. 2007). Thus, STPK-dependent phosphorylation might induce either positive or negative signaling to the FAS-II complexes in order to modulate mycolic acid biosynthesis and promote adaptation to environmental changes (Molle and Kremer 2010).

This view is supported by recent observations that in addition to KasA and KasB, *Mt*FabH is efficiently phosphorylated *in vitro* by several STPKs including PknH, PknA, and

PknF (Veyron-Churlet, Molle et al., 2009) and is also phosphorylated *in vivo* (Prisic, Dankawa et al. 2010; Chou, Wong et al. 2010). *Mt*FabH was phosphorylated exclusively on threonine residues, with Thr45 identified as the sole phosphoacceptor by mass spectrometry (Veyron-Churlet, Molle et al. 2009). The crystal structure of *Mt*FabH indicated that Thr45 was at the entrance to the substrate channel, suggesting that the phosphate group altered substrate accessibility and thus inhibited enzymatic activity (Scarsdale, Kanzanina et al. 2001). A phosphomimetic T45D mutant had significantly decreased enzymatic activity compared to a wild type or T45A mutant, and the T45A mutant completely abrogated phosphorylation by PknF and PknA (Veyron-Churlet, Molle et al. 2009). In this manner, altering the activities of several FAS II enzymes produces a tightly regulated system that allows for mycobacterial survival and adaptation under variable environmental conditions. While phosphorylation of each enzyme is accompanied by a partial reduction in activity, the cumulative effect could result in total cessation of mycolic acid production (Molle and Kremer 2010).

The pathogenicity of *Mtb* stems from its ability to alter the peptidoglycan, arabinogalactan, and mycolic components of its cell wall, and STPKs are involved in regulating and coordinating these processes. To better understand the mechanisms regulating STPKs and cell-wall architecture in *Mtb*, I biochemically and structurally characterized the extracellular domain of one STPK receptor kinase, PknH. Initial analysis of recombinantly produced PknH sensor domain by size-exclusion chromatography and small-angle x-ray scattering indicated the protein was soluble but completely unfolded. On-column refolding using decreasing amounts of guanidine hydrochloride produced refolded protein, which crystallized as a monomer with a completely novel protein fold made up of six alpha helices, seven beta strands and two disulfide bonds. The PknH sensor domain has a large and deep cleft with a mixed polar and hydrophobic surface. These results indicated PknH binds a small molecule ligand which may affect its quarternary structure and/or its localization.

In order to determine what ligand binds to the PknH sensor domain, I conducted native gel binding assays using purified *Mtb* arabinogalactan, lipoarabinomannan, and lipomannan. These assays indicated that *in vitro*, the PknH sensor domain does not bind to any of these purified cell wall components. Affinity chromatography with His-tagged PknH incubated with *Mtb* H37Rv cell lysates followed by intact mass, LC-MS, and ESI-MS analysis indicated that PknH does not bind glycolipids such as phosphoinositol or Ac_1PIM_2. Although the ligand for PknH was not identified, these studies set the stage for future researchers to identify the ligand and discover the function of PknH.

REFERENCES FOR CHAPTER 1

Alderwick, L.J., V. Molle, et al. (2006). Molecular structure of EmbR, a response element of Ser/Thr kinase signaling in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* **103**(8): 2558-2563.

Av-Gay, Y. and Everett, M. (2000). The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. *Trends Microbiol* **8**(5): 238-244.

Barthe, P., G.V. Mukamalova, et al. (2010). The structure of PknB extracellular PASTA domain from *Mycobacterium tuberculosis* suggests a ligand-dependent kinase activation. *Structure* **18**(5): 606-615.

Bhatt, A., N. Fujiwara, et al. (2007). Deletion of kasB in *Mycobacterium tuberculosis* causes loss of acid-fastness and subclinical latent tuberculosis in immunocompetent mice. *Proc Natl Acad Sci USA* **104**: 5157-5162.

Bhatt, A., V. Molle, et al. (2007). The *Mycobacterium tuberculosis* FAS-II condensing enzymes: their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. *Mol Microbiol* **64**(6): 1442-1454.

Bloch, K. and Vance, D. (1977). Control mechanisms in the synthesis of saturated fatty acids. *Annu Rev Biochem* **46**: 263-298.

Briken, V., S.A. Porcelli, et al. (2004). Mycobacterial lipoarabinomannan and related lipoglycans: from biogenesis to modulation of the immune response. *Mol Microbiol* **53**(2): 391-403.

Chao, J.D., D. Wong, et al. (2010). Protein kinase and phosphatase signaling in *Mycobacterium tuberculosis* physiology and pathogenesis. *Biochim Biophys Acta* **1804**(3): 620-627.

Choi, K.H. and L. Kremer, et al. (2000). Identification and substrate specificity of beta-ketoacyl (acyl carrier protein) synthase III mtFabH from *Mycobacterium tuberculosis*. *J Biol Chem* **275**: 28201-28207.

Cohen-Gonsaud, M., P. Barthe, et al. (2009). The *Mycobacterium tuberculosis* Ser/Thr kinase substrate Rv2175c is a DNA-binding protein regulated by phosphorylation. *J Biol Chem* **284**(29): 19290-19300.

Cole, S.T., R. Brosch, et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**(6685): 537-544.

Daffe, M. and Draper, P. (1998) The envelope layers of mycobacteria with reference to their pathogenicity. *Adv Microb Physiol* **39**: 131-203.

Dubnau, E., J. Chan, et al. (2000). Oxygenated mycolic acids are necessary for virulence of *Mycobacterium tuberculosis* in mice. *Mol Microbiol* **36**: 630-637.

Escuyer, V.E., M.A. Lety, et al. (2001). The role of the *embA* and *embB* gene products in the biosynthesis of the terminal hexaarabinofuranosyl motif of *Mycobacterium smegmatis* arabinogalactan. *J Biol Chem* **276**(52): 48854-48862.

Fernandes, N.D., and Kolattukudy, P.E. (1996). Cloning, sequencing and characterization of a fatty acid synthase-encoding gene from *Mycobacterium tuberculosis var. bovis BCG*. *Gene* **170**: 95-99.

Fulop, V. and Jones, D.T. (1999). Beta propellers: structural rigidity and functional diversity. *Curr Opin Struc Biol* **9**(6): 715-721.

Gay, L.M. (2007). Structural studies of PknE and an FHA domain from Rv1747: Signal Transduction Molecules in *Mycobacterium tuberculosis*. Ph.D. Dissertation for MCB Department, UC Berkeley.

Gay, L.M., H.L. Ng, et al. (2006). A conserved dimer and global conformational changes in the structure of apo-PknE Ser/Thr protein kinase from *Mycobacterium tuberculosis*. *J Mol Biol* **360**(2): 409-420.

Gee, C.L., K.G. Papavinasasundaram, et al. (2012). A phosphorylated pseudokinase complex controls cell-wall synthesis in mycobacteria. *Sci Signal* **5**(208): ra7.

Glickman, M.S., and Jacobs, W.R., Jr. (2001). Microbial pathogenesis of *Mycobacterium tuberculosis*: dawn of a discipline. *Cell* **104**: 477-485.

Good, M.C., A.E. Greenstein, et al. (2004). Sensor domain of the *Mycobacterium tuberculosis* receptor Ser/Thr protein kinase, PknD, forms a highly symmetric beta propeller. *J Mol Biol* **339**(2): 459-469.

Grebe, T.W. and Stock, J.B. (1999). The histidine protein kinase superfamily. *Adv Microbial Physiol* **41**: 139-227.

Greenstein, A.E., J.A. MacGurn, et al. (2007). *M. tuberculosis* Ser/Thr protein kinase D phosphorylates an anti-anti-sigma factor homolog. *PLoS Pathog* **3**(4): e49.

Hanks, S. K. and Hunter, T. (1995). The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* **9**(8): 576-596.

Hoch, J.A. (2000). Two-component and phosphorelay signal transduction. *Curr Opin Microbiol* **3**(2): 165-170.

Jayakumar, D., W.R. Jacobs, Jr., et al. (2008). Protein kinase E of *Mycobacterium tuberculosis* has a role in the nitric oxide stress response and apoptosis in a human macrophage model of infection. *Cell Microbiol* **10**(2): 365-374.

Kang, C.M., D.W. Abbott, et al. (2005). The *Mycobacterium tuberculosis* serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape. *Genes Dev* **19**(14): 1692-1704.

Kang, C.M., S. Nyayapathy, et al. (2008). Wag31, a homologue of the cell division protein DivIVA, regulates growth, morphology and polar cell wall synthesis in mycobacteria. *Microbiol* **154**(Pt 3): 725-735.

Kannan, N., S.S. Taylor, et al. (2007). Structural and functional diversity of the microbial kinome. *PLoS Biol* **5**(3): e17.

Kremer, L., J.D. Douglas, et al. (2000). Thiolactomycin and related analogues as novel anti-mycobacterial agents targeting KasA and KasB condensing enzymes in *Mycobacterium tuberculosis*. *J Biol Chem* **275**(22): 16857:16864.

Kremer, L., L.G. Dover, et al. (2002). Mycolic acid biosynthesis and enzymatic characterization of the beta-ketoacyl-ACP synthase A-condensing enzyme from *Mycobacterium tuberculosis*. *Biochem J* **364**: 423-430.

Kornev, A.P. and Taylor, S.S. (2010). Defining the conserved internal architecture of a protein kinase. *Biochim Biophys Acta* **1804**(3): 440-444.

Kumar, D. and Narayanan, S. (2012). *pknE*, a serine/threonine kinase of *Mycobacterium tuberculosis* modulates multiple apoptotic paradigms. *Infect Genet Evol* **12**(4): 737-747.

Letek, M., E. Ordóñez, et al. (2008) DivIVA is required for polar growth in the MreB-lacking rod-shaped actinomycete *Corynebacterium glutamicum*. *J Bacteriol* **190**(9): 3283-3292.

Manning, G., D. B. Whyte, et al. (2002). The protein kinase complement of the human genome. *Science* **298**(5600): 1912-1934.

Marrakchi, H., F. Bardou, et al. (2008) *Chapter 4: A Comprehensive Overview of Mycolic Acid Structure and Biosynthesis*. The Mycobacterial Cell Envelope (ed. M. Daffe and J.-M. Reyrat) ASM Press, Washington, DC.

Mir, M., J. Asong, et al. (2011). The extracytoplasmic domain of the *Mycobacterium tuberculosis* Ser/Thr kinase PknB binds specific muropeptides and is required for PknB localization. *PLoS Pathog* **7**(7): e1002182.

Molle, V., A.K. Brown, et al. (2006). The condensing activities of the *Mycobacterium tuberculosis* type II fatty acid synthase are differentially regulated by phosphorylation. *J Biol Chem* **281**: 30094-30103.

Molle, V., C. Girard-Blanc, et al. (2003). Protein PknE, a novel transmembrane eukaryotic-like serine/threonine kinase from *Mycobacterium tuberculosis*. *Biochem Biophys Res Commun* **308**(4): 820-825.

Molle, V. and Kremer, L. (2010). Division and cell envelope regulation by Ser/Thr phosphorylation: Mycobacterium shows the way. *Mol Microbiol* **75**(5): 1064-1077.

Molle, V., L. Kremer, et al. (2003). An FHA phosphoprotein recognition domain mediates protein EmbR phosphorylation by PknH, a Ser/Thr protein kinase from *Mycobacterium tuberculosis*. *Biochemistry* **42**(51): 15300-15309.

Narayan, A., P. Sachdeva, et al. (2007). Serine threonine protein kinases of mycobacterial genus: phylogeny to function. *Physiol Genomics* **29**(1): 66-75.

Nguyen, L., N. Scherr, et al. (2007). Antigen 84, an effector of pleiomorphism in *Mycobacterium smegmatis.* *J Bacteriol* **189**(21): 7896-7910.

Ortiz-Lombardia, M., F. Pompeo, et al. (2003). Crystal structure of the catalytic domain of the PknB serine/threonine kinase from *Mycobacterium tuberculosis.* *J Biol Chem* **278**(15): 13094-13100.

Papavinasasundaram, K. G., B. Chan, et al. (2005). Deletion of the *Mycobacterium tuberculosis pknH* gene confers a higher bacillary load during the chronic phase of infection in BALB/c mice. *J Bacteriol* **187**(16): 5751-5760.

Parish, T., D. A. Smith, et al. (2003). Deletion of two-component regulatory systems increases the virulence of *Mycobacterium tuberculosis*. *Infect Immun* **71**(3): 1134-1140.

Pereira, S.F.F., L. Goss, et al. (2011) Eukaryote-Like Serine Threonine Kinases and Phosphatases in Bacteria. *Microbiol. Mol. Biol. Rev.* **75**(1): 192-212.

Pons, T., R. Gomez, et al. (2003). Beta-propellers: associated functions and their role in human diseases. *Curr Med Chem* **10**(6): 505-524.

Portevin, D., C. De Sousa-D'Auria, et al. (2004). A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. *Proc Natl Acad Sci USA* **101**: 314-319.

Prisic, S., S. Dankwa, et al. (2010). Extensive phosphorylation with overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein kinases. *Proc Natl Acad Sci USA* **107**(16): 7521-7526.

Scarsdale, J. N., G. Kazanina, et al. (2001). Crystal Structure of the *Mycobacterium tuberculosis* β-Ketoacyl-Acyl Carrier Protein Synthase III. *J Biol Chem* **276**: 20516-20522.

Shaeffer, M.L., G. Agnihotri, et al. (2001). Purification and biochemical characterization of the *Mycobacterium tuberculosis* beta-keto-acyl carrier protein synthases KasA and KasB. *J Biol Chem* **276**: 20516-20522.

Shah, I.M., M.H. Laaberki, et al. (2008). A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. *Cell* **135**(3): 486-496.

Sharma, K., H. Chandra, et al. (2004). PknH, a transmembrane Hank's type serine/threonine kinase from *Mycobacterium tuberculosis* is differentially expressed under stress conditions. *FEMS Microbiol Lett* **233**(1): 107-113.

Sharma, K., M. Gupta, et al. (2006a). EmbR, a regulatory protein with ATPase activity, is a substrate of multiple serine/threonine kinases and phosphatase in *Mycobacterium tuberculosis*. *FEBS J* **273**(12): 2711-2721.

Sharma, K., M. Gupta, et al. (2006b). Transcriptional control of the mycobacterial embCAB operon by PknH through a regulatory protein, EmbR, *in vivo*. *J Bacteriol* **188**(8): 2936-2944.

Stock, J.B., A.J. Ninfa, et al. (1989). Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol Rev* **53**(4): 450-90.

Thakur, M. and Chakraborti, P.K. (2008). Ability of PknA, a mycobacterial eukaryotic-type serine/threonine kinase, to transphosphorylate MurD, a ligase involved in the process of peptidoglycan biosynthesis. *Biochem J* **415**(1): 27-33.

Veyron-Churlet, R., V. Molle, et al. (2009). The *Mycobacterium tuberculosis* beta-ketoacyl-acyl carrier protein synthase III activity is inhibited by phosphorylation on a single threonine residue. *J Biol Chem* **284**: 6414-6424.

Wietzorrek, A., and Bibb, M. (1997). A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. *Mol Microbiol* 25(6):1181-4.

Yuan, Y., Y. Zhu, et al. (1998). The effect of oxygenated mycolic acids composition on cell wall function and macrophage growth in *Mycobacterium tuberculosis*. *Mol Microbiol* **29**: 1449-1458.

Zhang, N., J.B. Torrelles, et al. (2003). The Emb proteins of mycobacteria direct arabinosylation of lipoarabinomannan and arabinogalactan via an N-terminal recognition region and a C-terminal synthetic region. *Mol Microbiol* **50**(1): 69-76.

CHAPTER 2: PRODUCING A SOLUBLE, FOLDED PKNH SENSOR DOMAIN

I. Introduction

In order to survive and persist inside human macrophage cells, *Mtb* utilizes an array of 11 serine/threonine protein kinases, designated PknA through PknL, to sense and respond to environmental signals (Alber 2009; Figure 1.2). Two of these STPKs, PknB and PknD, have extracellular domains (ECDs) that are known to be folded (Barthe, Mukamolova et al. 2010; Good, Greenstein et al. 2004). The PknD sensor domain forms a rigid, six-bladed beta-propeller (Good, Greenstein et al. 2004), while the PknB sensor domain has four PASTA domains (Barthe, Mukamolova et al. 2010) that bind peptidoglycan fragments and localize the kinase to sites of peptidoglycan turnover to regulate cell growth and division (Shah, Laaberki et al. 2008; Mir, Asong et al. 2010).

Two other STPKs, PknE and PknH, are predicted to have folded ECDs (Gay 2007; Baer 2010). The PknE ECD is predicted to have a protein disulfide isomerase-like fold similar to *E. coli* DsbG (Gay, 2007), while bioinformatic analysis of the PknH ECD via BLAST (Altschul, Madden et al. 1997) shows it has a 36% identity to LppH, an uncharacterized mycobacterial lipoprotein (expect score of $2e^{-28}$, 67 identities over 188 residues). With the exception of PknB, the roles of the PknD, PknE, and PknH extracellular domains in kinase regulation, including their activating ligands, remain unknown. These sensor domains may regulate STPK activity by directing kinase localization and/or oligomerization (Mir, Asong, et al. 2010; Palova, Hercik, et al. 2007). In order to determine the function of the PknH sensor domain, I set out to produce it in a soluble and properly folded form. This chapter details how this goal was accomplished.

II. Materials and Methods

A. Cloning, Expression, and Purification

PknH gene fragments were amplified from genomic *Mycobacterium tuberculosis* DNA (strain H37Rv) obtained from the Mycobacteria Research Laboratories at Colorado State University**.** The different versions of the PknH sensor domain were cloned into pDONR221 donor vectors and recombined with Gateway expression vectors pHGWA and pHMGWA. All constructs were sequenced to confirm proper insertion and orientation. The tags include an N-terminal His_6 and N-terminal His_6-maltose binding protein (His_6-MBP) tag, respectively. Each tag is followed by a tobacco etch virus protease site (Busso, Delagoute-Busso et al. 2005) which leaves behind a Gly-His fragment after tag removal.

Proteins were produced in *E. coli* BL21 CodonPlus cells (Stratagene) using auto-induction media (Studier 2005). Cultures (5 mL) were grown at 37 °C for six hours and shifted to 24 °C for 17-18 hours at 180 rpm. Cell pellets were harvested by centrifugation and lysed by sonication, with the resulting resuspension clarified by a high-speed centrifugation at 14000 rpm for 20 minutes. The clarified supernatant was filtered through a 0.22 um syringe filter. 16 uL samples of each fraction (lysate, supernatant, filtered supernatant, and pellet) were combined with 4 uL of 5X reducing SDS dye. 10 uL of each sample was applied to a 12% Tris-Glycine pre-cast gel for SDS-PAGE analysis

For typical large –scale purifications, 4 L of *E. coli* BL21 (DE3) CodonPlus cells were grown at 37°C for six hours and 30 °C for 18 hours in auto-induction media (Studier 2005).

Pellets were harvested by centrifugation for 15 minutes at 4500 rpm in an Avanti centrifuge. Each pellet was resuspended in Ni-IMAC Buffer A (300 mM NaCl, 25 mM HEPES pH 8.0, 5% glycerol, 25 mM imidazole and 250 uM AEBSF protease inhibitor). The cells were lysed by sonication and centrifuged at 16000 rpm for 30 minutes. The clarified supernatant was loaded onto a pre-packed 5 mL Ni-NTA sepharose column that had been equilibrated with Buffer A. Bound protein was eluted with Buffer B (Buffer A plus 300 mM imidazole). The eluted protein was dialyzed against 4 L of Buffer C (Buffer A minus imidazole) overnight and His_6-tagged TEV protease was added in a 1:100 ratio of protease to protein. The dialysis step and simultaneous TEV cleavage was allowed to proceed overnight at 4 °C.

To remove the His_6-MBP tag, the dialyzed protein was run over a second 5 mL Ni-NTA Sepharose column and the cleaved protein was collected in the flow-through fractions (Figure 2.3). The protein was further purified using size-exclusion chromatography (SEC) with a HiLoad 26/60 Superdex S75 column run in Buffer D (100 mM NaCl, 25 mM HEPES pH 8.0 and 5% glycerol) (Figure 2.3). The purified protein eluted as a monomer and was concentrated to 5 mg/mL for crystallization trials using a Mosquito Crystallization Robot.

B. Small-Angle X-Ray Scattering (SAXS)

His_6-MBP tagged PknH sensor domain (residues 438-624) was purified as described in Section IIB above and concentrated to 1.8 mg/mL and 1.2 mg/mL in SAXS buffer (100 mM NaCl, 25 mM HEPES pH 8.0 and 5% glycerol). SAXS data were collected at the SIBYLS beamline at the Lawrence Berkeley National Laboratory Advanced Light Source using a wavelength, λ, of 1.03 Å with the sample-to-detector distance set to 1.5 m, resulting in scattering vectors, q, ranging from 0.01 $Å^{-1}$ to 0.30 $Å^{-1}$. The data were acquired at 20 °C, and short and long time exposures (0.5 s, 5 s) were merged for the calculations using the entire scattering profile. The experimental SAXS data were measured at 1.8 mg/mL and 1.2 mg/mL. Data were processed as described (Huga 2009). The radius of gyration R_G was derived by the Guinier approximation $I(q) = I(0) \exp(-q^2 R_G^2/3)$ with the limits $qR_G < 1.3$. The interference-free SAXS profile was estimated by extrapolating the measured scattering curves to infinite dilution (Figure 2.4).

C. Inclusion –body purification and refolding

PknH sensor domain (residues 438-624) with an N-terminal His_6-tag was expressed in BL21 (DE3) CodonPlus cells (Stratagene) using auto-induction media (Studier 2005). This smaller, His_6-tagged protein went primarily to the pellet (Figure 2.5), presumably due to its presence in inclusion bodies.

Cells were grown at 37 °C for six hours and then at 30 °C for 18 hours. Cell paste was resuspended in Buffer E (200 mM NaCl, 5 % v/v glycerol and 20 mM Bis-Tris propane at pH 8.5) plus protease inhibitors E-64 and leupeptin, and lysed by sonication on ice. The lysate was centrifuged for 40 minutes at 16000 rpm in a Sorvall centrifuge with an SS-34 rotor. The resulting supernatant was discarded, and the pellet was resuspended in Buffer A plus 0.1% Triton X-100. This mixture was sonicated and centrifuged at 9000 rpm for 20 minutes. This supernatant was discarded and the detergent extraction/sonication/centrifugation step was repeated twice, for a total of 3 detergent extractions.

The pellet was resuspended at 4 °C overnight in Buffer E plus 6 M guanidinium hydrochloride, 1 mM reduced glutathione and 0.1 mM oxidized glutathione. The solution was sonicated and centrifuged at 17000 rpm for 30 minutes. The supernatant was incubated with Ni-NTA agarose beads (Qiagen) for 4 hours. Step-wise washes (100 mL for every 4 mL of Ni-NTA resin) that reduced the GuHCl concentration by 1 M in each step were used to gradually remove guanidinium hydrochloride and allow the protein to refold in the presence of the 10:1 ratio of reduced to oxidized glutathione. The protein was eluted with Buffer E plus 300 mM imidazole. After dialysis against 4 L of Buffer E to remove the imidazole and glutathione, the protein was incubated with TEV protease overnight at 4°C. Cleaved PknH protein was purified by a second IMAC step and further purified by size exclusion chromatography using a HiLoad 26/60 Superdex 75 column in Buffer F (100 mM NaCl, 5% v/v glycerol and 20 mM PIPES pH 6.5). The protein eluted as a monomer off of the size-exclusion column.

D. Circular Dichroism

A sample of purified, refolded PknH sensor domain (residues 438-624) was extensively dialyzed into 100 mM NaF at pH 6.5 and concentrated to 0.67 mg/mL. Circular dichroism (CD) spectra were recorded on an Aviv 410 spectropolarimeter with a Peltier temperature-controlled cell holder. For all samples, each CD measurement is an average of the signal at 222 nm over 1 min in a 1 cm path length cuvette. Data were collected at 25 °C from 250 to 200 nm in 1-nm steps, at a scanning speed of 200 nm/min. The results are reported as mean residue ellipticity (degrees cm^2 $dmol^{-1}$).

E. Limited Proteolysis

N-terminally His_6-tagged PknH sensor domain protein was purified and refolded as described in Section C and concentrated to 1 mg/mL. 200 ug of PknH was incubated with 200 ng of thermolysin (Sigma) on ice and divided equally into two 1.5 mL tubes. The reducing agent TCEP was added to one of the tubes to a final concentration of 5 mM. 12 uL aliquots of the PknH/thermolysin proteolysis mix were removed at 0, 2, 5, 10, 20, 30 and 60 minutes while one set of aliquots sat overnight at 4 °C. To stop the proteolytic reaction, aliquots were removed from the eppendorf tube at the appropriate times and incubated with 1 uL of 2% trifluoroacetic acid (TFA) and 3 uL of 5X SDS loading dye and run on a 12% Tris-Glycine gel.

III. Results and Discussion

Bioinformatic analysis using the program TMHMM v2.0 (Krogh and Larsson 2001) showed that *Mtb* H37Rv PknH, which is 626 amino acids long, was predicted to have a transmembrane helix from position 404 to 426. The first predicted residue in the extracellular sensor domain is Arg 427. The start and end points for the PknH sensor domain were chosen based on predicted secondary structural features as determined by the ROBETTA BETA web server (Kim, Chivian et al. 2004). Several different N-terminal start sites for the protein were chosen and combined with one of two end sites at the C-terminus of the sensor domain (Figure 2.1).

Small-scale expression tests were conducted to determine which constructs were highly expressed and soluble (Figures 2.2 and 2.3) when combined with an N-terminal His_6-MBP

purification tag. The longer PknH constructs were better expressed than the shorter ones (Figure 2.2), and construct spanning residues 438-624 with an N-terminal His_6-MBP tag was chosen for further purification efforts.

This PknH sensor domain construct could be readily produced and purified via two immobilized metal affinity chromatography steps and a final size-exclusion chromatography step after tag removal (Figure 2.3). However, analysis of the protein via small-angle X-ray scattering indicated it was unfolded, as evidenced by the observed decrease in scattering intensity at small angles in the Guinier plot (Figure 2.4); this observed decrease is caused by significant inter-particle repulsion present when a protein is unfolded (Mertens, Svergun 2010). In contrast, a protein that was aggregated instead of unfolded would have produced a Guinier plot with a sharp increase in intensity at small scattering angles (Mertens, Svergun 2010). Because the PknH sensor domain contains four conserved Cys residues, I reasoned that disulfide bonds may stabilize the structure. In this case, the protein was likely unfolded due to its exposure to the reductive *E. coli* cytoplasm. To alleviate this problem, I developed an on-column refolding protocol involving stepwise reductions in the levels of the denaturant guanidinium hydrochloride in the presence of a 10:1 ratio of reduced to oxidized glutathione.

Circular dichroism and limited proteolysis experiments on refolded His_6-tagged PknH demonstrated the refolding was successful and the sensor domain was now properly folded. The curve in Figure 2.6 displays a negative molar ellipticity at 208 nm that is characteristic of proteins with some beta sheet structure, while the upward slope of the curve towards positive mean residue ellipticity near the 198 nm wavelength suggests the presence of alpha-helices (Correa and Ramos 2009). To further confirm that PknH was properly folded, I subjected the protein to limited proteolysis with and without the addition of 5 mM TCEP (Figure 2.7). As expected for a protein with disulfide bonds, PknH was more sensitive to thermolysin in the presence of TCEP and was cleaved into several fragments; the PknH that was not treated with TCEP was more resistant to thermolysin (Figure 2.7). These experiments represent a critical first step in solving the three-dimensional structure of the PknH sensor domain and identifying its ligands and functions.

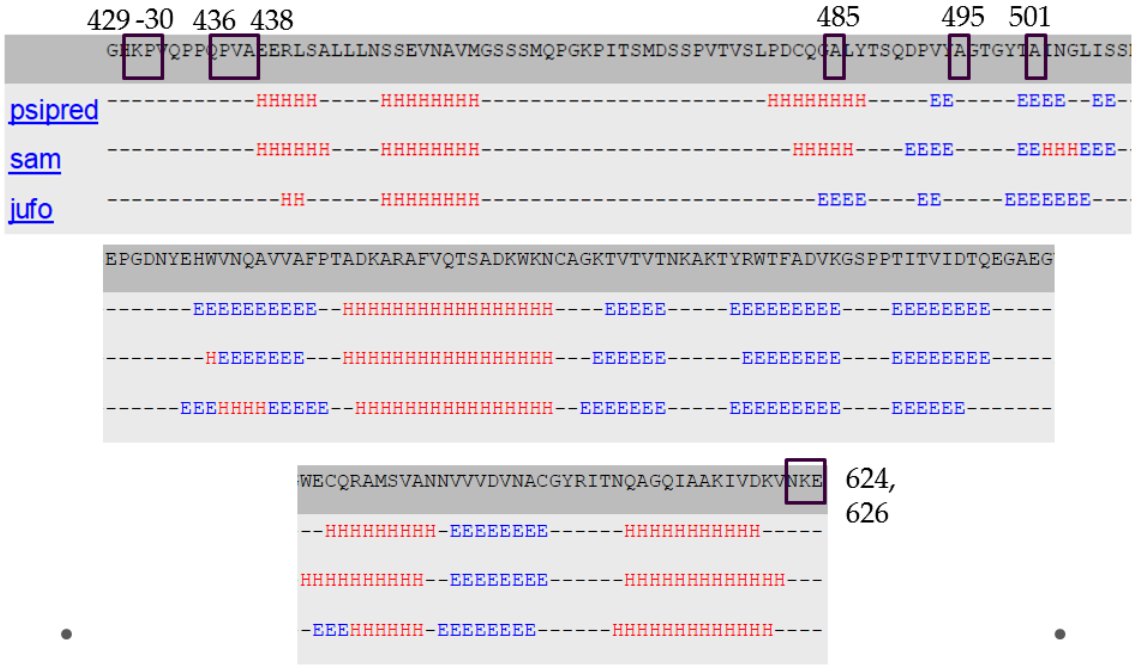


Figure 2.1 Secondary structure predictions from the ROBETTA BETA server for the PknH sensor domain. "H" is for alpha helix, and "E" is for beta strand. The PknH sensor domain extends from position 427 to 626, and boxed residues with numbers indicate the start and end sites chosen for cloning into Gateway-based vectors.

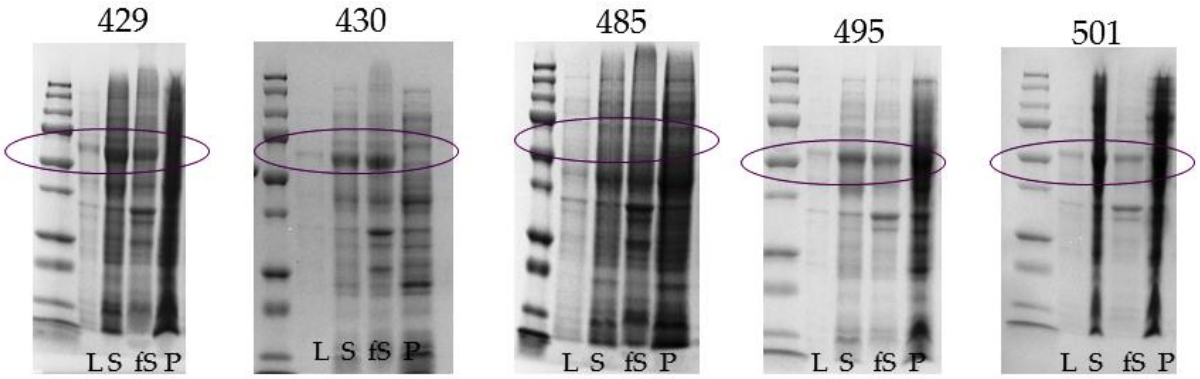


Figure 2.2 Expression test results for PknH sensor domain constructs of varying lengths. SeeBlue Plus 2 protein ladder samples are in leftmost lane of each gel. The numbers above the gels represent the start site for the construct, and all constructs shown here ended at position 624. "L" is for the cell lysate fraction, "S" is for the supernatant fraction, "fS" is for supernatant filtered through a 0.22 um syringe filter, and "P" is for the pellet fraction. Purple ovals indicate the expected location for the protein bands of the PknH construct tested. No expression was detected for the PknH sensor domain construct that began at residue 485.

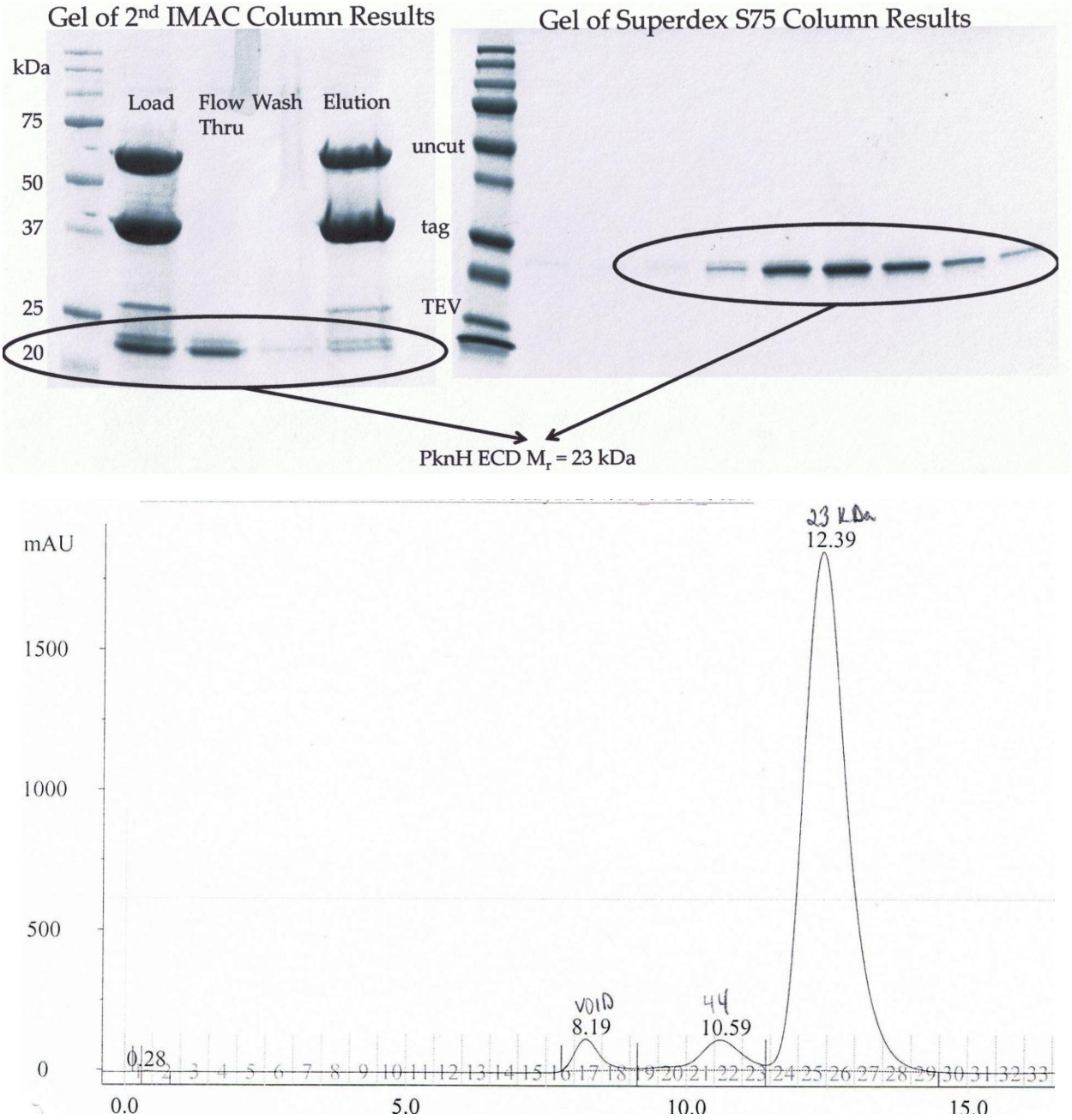


Figure 2.3 SDS-PAGE of PknH sensor domain (residues 438-624) during the 2nd IMAC purification (left gel) and after final purification via size-exclusion chromatography (right gel). The gel on the left shows the state of the protein as it was loaded onto the second Ni-NTA column (Load), what flowed through the column (Flow Thru), what was washed with 25 mM imidazole (Wash), and what eluted from the column with 300 mM imidazole (Elution). The right gel shows fractions 23-29 from an analytical sizing column run (bottom panel); the purified protein eluted as a monomer at 12.39 mL.

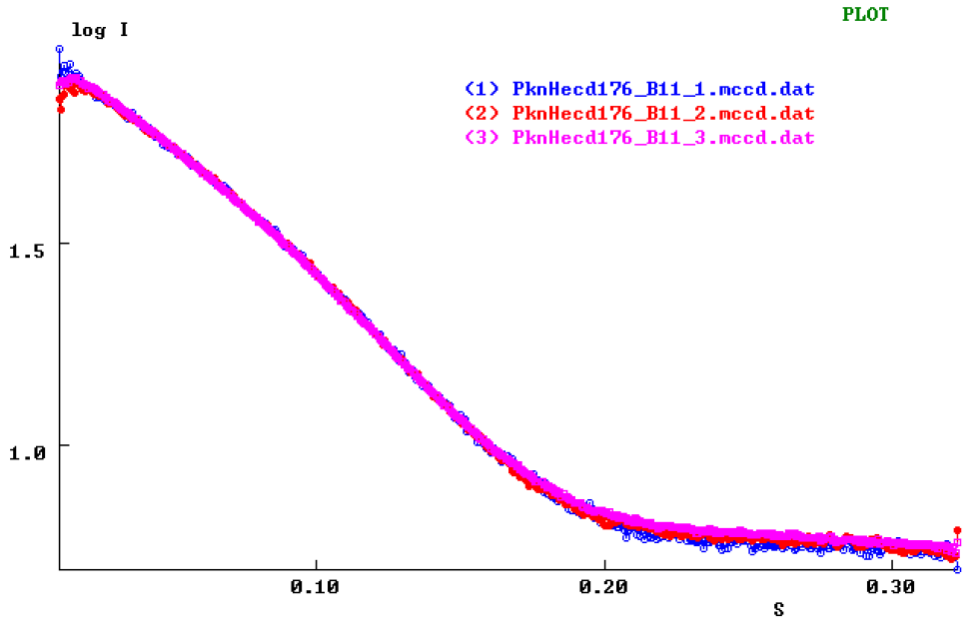


Figure 2.4 SAXS scattering profile of the refolded PknH sensor domain at 1.8 mg/mL. Three replicates were performed and corresponding curves are shown in blue, red, and magenta.

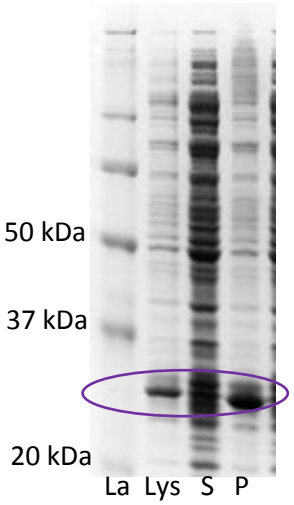


Figure 2.5 SDS-PAGE of N-terminally His6-tagged PknH sensor domain. The PrecisionPlus molecular standards ladder is labeled "La", while "Lys" indicates the lysate, "S" is for supernatant, and "P" is for pellet. The purple oval indicates the bands corresponding to the PknH sensor domain at the expected molecular mass of 24 kD. The larger, darker size of the band in the pellet lane indicated this PknH sensor domain fragment was expressed in inclusion bodies.

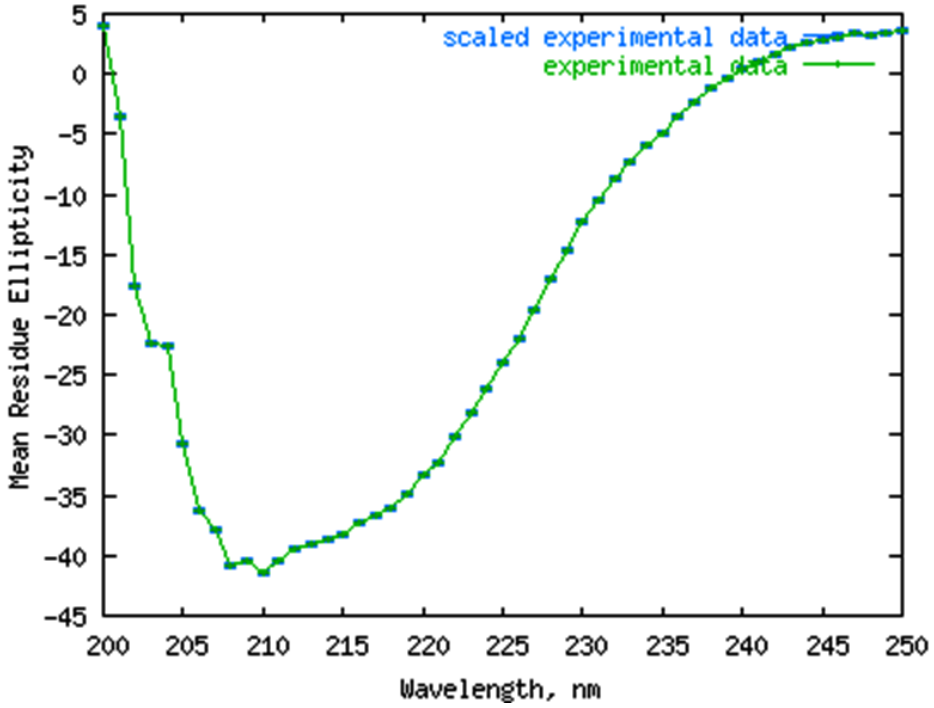


Figure 2.6. Far UV CD spectrum of the N-terminally His_6-tagged PknH sensor domain after refolding.

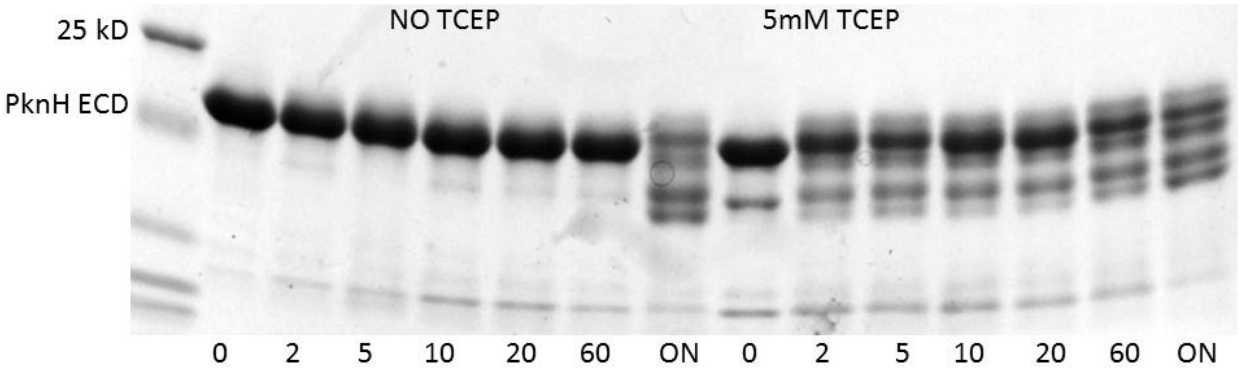


Figure 2.7 Limited proteolysis of refolded N-terminally His_6-tagged PknH sensor domain after incubation with the protease thermolysin. The first lane of this 12% Tris-Glycine gel is the PrecisionPlus ladder, while lanes 2-7 and 9-14 indicate the number of minutes that PknH was incubated with protease at 4 °C. "ON" denotes an overnight incubation with protease. The samples on the left side of the gel (Lanes 2-8) did not have any reducing agent, while the samples on the right (Lanes 9-15) had 5 mM TCEP added. All gel samples had a 1:1000 ratio of PknH sensor domain: thermolysin.

REFERENCES FOR CHAPTER 2

Alber, T. (2009). Signaling mechanisms of the *Mycobacterium tuberculosis* receptor Ser/Thr protein kinases. *Curr Op Struct Biol* **19**: 650-657.

Altschul, S.F., T.L. Madden, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.

Baer, C.E. (2010). Mechanisms of *Mycobacterium tuberculosis* Serine/Threonine Protein Kinase Activation. Ph.D. Dissertation for Biophysics Department, UC Berkeley.

Barthe, P., G.V. Mukamolova, et al. (2010). The structure of PknB Extracellular PASTA Domain from *Mycobacterium tuberculosis* Suggests a Ligand-Dependent Kinase Activation. *Struct* **18**: 606-615.

Busso, D., B. Delagoutte-Busso, et al. (2005). Construction of a set of Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*. *Anal Biochem* **343**(2): 313-321.

Correa, D.H.A and Ramos, C. (2009). The use of circular dichroism spectroscopy to study protein folding, form and function. *African J Biochem Res* **3**(5): 164-173.

Gay, L.M. (2007). Structural studies of PknE and an FHA domain from Rv1747: Signal Transduction Molecules in *Mycobacterium tuberculosis*. Ph.D. Dissertation for MCB Department, UC Berkeley.

Good, M.C., A.E. Greenstein, et al. (2004). Sensor domain of the *Mycobacterium tuberculosis* receptor Ser/Thr protein kinase, PknD, forms a highly symmetric beta propeller. *J Mol Biol* **339**(2): 459-469.

Kim, D.E., D. Chivian, et al. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32** (Suppl 2): 526-531.

Krogh, A., B. Larsson, et al. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**(3): 567-580.

Mertens, H.D., and Svergun, D.I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J Struct Biol* **172**(1): 128-141.

Mir, M., J. Asong, et al. (2011). The extracytoplasmic domain of the *Mycobacterium tuberculosis* Ser/Thr kinase PknB binds specific muropeptides and is required for PknB localization. *PLoS Pathog* **7**(7): e1002182.

Molle, V. and Kremer, L. (2010). Division and cell envelope regulation by Ser/Thr phosphorylation: Mycobacterium shows the way. *Mol Microbiol* **75**(5): 1064-1077.

Pallova, P., K. Hercik, et al. (2007). A eukaryotic-type serine/threonine protein kinase StkP of *Streptococcus pneumonia* acts as a dimer *in vivo*. *Biochem Biophys Res Commun* **355**: 526-530.

Shah, I.M., M.H. Laaberki, et al. (2008). A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. *Cell* **135**(3): 486-496.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**(1): 207-234.

CHAPTER 3: SENSOR DOMAIN OF THE *MYCOBACTERIUM TUBERCULOSIS* PKNH RECEPTOR SER/THR PROTEIN KINASE ADOPTS A NOVEL FOLD

I. Introduction

Mycobacterium tuberculosis (*Mtb*) is a persistent human pathogen that currently infects one-third of the world's population and causes over 1.7 million deaths per year (WHO 2010). The pathogenicity of *Mtb* stems from its ability to alter its developmental programs and metabolism in different host niches. After phagocytosis by alveolar macrophages, *Mtb* slows growth and alters the composition of cell wall mycolic and fatty acids to survive the nutrient poor phagocytic environment and resist microbicides such as nitric oxide and reactive oxygen species (Schnappinger, Ehrt et al. 2003). However, little is known about the developmental programs and molecular signals that trigger these adaptive responses. Candidate sensor molecules for transmitting environmental signals into adaptive responses include the 11 eukaryotic-like Ser/Thr protein kinases (STPKs) encoded in the *Mtb* genome, nine of which have an intracellular N-terminal kinase domain linked via a single transmembrane helix to an extracellular C-terminal sensor domain (Av-Gay and Everett 2000). Recent sequencing projects indicate eukaryotic-like STPKs are ubiquitous and exist in many prokaryotes, including a wide range of pathogenic bacteria (Galperin, Higdon et al. 2010). Since their discovery, STPKs have been shown to regulate diverse cellular functions, such as exit from dormancy (Shah, Laaberki et al. 2008; Shah, Dworkin et al. 2010) protein secretion (Mougous, Gifford et al. 2007), cell division (Fiuza, Canova et al. 2008), sporulation (Madec, Laskiewicz et al. 2002; Madec, Stensballe et al. 2003) and cell-wall biosynthesis (Fiuza, Canova et al. 2008).

The first bacterial STPK kinase domain (KD) structures, which revealed nucleotide complexes of the *Mtb* PknB KD, demonstrated that bacterial and eukaryotic STPKs share close structural similarities and common modes of substrate recognition and regulation (Young, Delagoutte et al. 2003; Ortiz-Lombardia, Pompeo et al. 2003). Despite advances in understanding the kinase domains of STPKs, only two of the *Mtb* STPK sensor domains have been structurally characterized. The PknD sensor domain structure was found to form a rigid, six-bladed beta-propeller with a flexible linker to the transmembrane helix (Good, Greenstein et al. 2004), while the PknB sensor domain was found to have four PASTA domains (Barthe, Mukamolova et al. 2010) that bind peptidoglycan fragments and localize the kinase to sites of peptidoglycan turnover to regulate cell growth and division (Shah, Laaberki et al. 2008; Mir, Asong et al. 2010). To further our understanding of STPK receptor signaling, I now describe the x-ray crystal structure of the extracellular sensor domain of the *Mtb* STPK PknH.

II. MATERIALS AND METHODS

A. Protein production and structure determination

To characterize the PknH sensor, I expressed the extracellular domain (ECD; residues 435-626) beginning eight residues after the predicted transmembrane helix. PknH residues 435-626 were PCR-amplified from *Mycobacterium tuberculosis* H37Rv genomic DNA and cloned into the pET28-based destination vector pHGWA using Gateway enzymes (Invitrogen). The pHGWA vector has an N-terminal His_6 tag followed by a TEV protease cleavage site (Busso,

Delagoutte-Busso et al. 2005) that leaves Gly-His at the N-terminus of the protein after TEV cleavage. The protein was expressed in BL21 (DE3) CodonPlus cells (Stratagene) using auto-induction media (Studier 2005). Cells were grown at 37 °C for six hours and then at 30 °C for 18 hours. Cell paste was resuspended in Buffer A (200 mM NaCl, 5% v/v glycerol and 20 mM Bis-Tris propane at pH 8.5) plus protease inhibitors E-64 and leupeptin, and lysed by sonication on ice. The lysate was centrifuged for 40 minutes at 16000 rpm and the resulting supernatant was discarded. The pellet was resuspended in Buffer A plus 0.1% Triton X-100 for 6 – 8 hours at 4 °C. This mixture was sonicated and centrifuged at 9000 rpm for 20 minutes. The supernatant was discarded and the detergent extraction step was repeated twice. The pellet was resuspended at 4 °C overnight in Buffer A plus 6 M guanidine hydrochloride (GuHCl), 1 mM reduced glutathione, and 0.1 mM oxidized glutathione. The GuHCl solution was sonicated and centrifuged at 17000 rpm for 30 minutes. The supernatant was incubated with Ni-NTA agarose beads (Qiagen) for 4 hours, and step-wise washes that reduced the GuHCl concentration in 1 M increments were used to refold the protein in the presence of the 10:1 ratio of reduced to oxidized glutathione. The protein was eluted with Buffer A plus 300 mM imidazole. After dialysis into Buffer A to remove the imidazole, the protein was incubated with TEV protease overnight at 4 °C. Cleaved PknH protein was purified by a second IMAC step and further purified by size exclusion chromatography using a HiLoad 26/60 Superdex 75 column in Buffer B (100 mM NaCl, 5% v/v glycerol and 20 mM PIPES pH 6.5).

A single peak corresponding to the PknH sensor domain monomer was concentrated to 5 mg/mL and was crystallized at 4 °C by vapor diffusion in a 1:1.5 ratio of protein to 0.1 M Bis-Tris (pH 5.5), 0.2 M ammonium acetate, 25% PEG 3350. Additive screening of the initial crystal hit led to improved crystals with terbium nitrate, which were used for SAD. Optimal terbium crystals were grown overnight at 18 °C using 250 nL of crystallization buffer, 500 nL of 5 mg/mL PknH in Buffer B, and 100 nL of 100 mM terbium nitrate. Crystals were cryoprotected with mother liquor containing 15% v/v glycerol.

The crystal structure was determined at 1.7-Å resolution by single-wavelength anomalous diffraction (SAD) analysis of a terbium derivative (Table 1). The entire sequence from residues 435 to 626 was ordered. Data collection on the cryo-cooled crystal was performed on Beamline 8.3.1 at the Lawrence Berkeley National Laboratory Advanced Light Source (MacDowell et al. 2004). Data were reduced and scaled with HKL2000 (Otwinoski and Minor 1997). The structure was determined using PHENIX (Adams, Afonine et al. 2011) and the model was adjusted manually using Coot (Emsley and Cowtan 2004). Phenix.autosol found two terbium sites per asymmetric unit, and phenix.autobuild produced a model with 191 residues and an R_{free} of 24% after seven cycles of automatic building and refinement. Building and refinement were completed with phenix.refine and Coot and included addition of a single ordered molecule of BIS-TRIS buffer that coordinated one of the two terbium atoms. The final model was validated using MolProbity (Chen, Arendall et al. 2010).

III. RESULTS AND DISCUSSION

A. PknH sensor domain structure

The PknH sensor domain contains six alpha helices and seven anti-parallel beta strands with α1-α2-α3-α4-β1-β2-α5-β3-β4-β5-β6-β7-α6 topology (Figure 1a). Two intramolecular disulfide bonds link α3 to α5 (C482 – C545) and β6 to β7 (C587 – C604). A 22-residue irregular

loop connects α2 and α3. The most prominent feature is a large v-shaped central cleft (Figure 1b). Five of the seven anti-parallel beta-strands (β1/β2 and β5 – β7) make up one side of this cleft, while alpha helices α3 to α5 and beta strands β3 and β4 comprise the other side. The α2-α3 loop forms the rim of the cleft, and residues 486-490 in the α3-α4 loop line the cleft inner wall (Figure 1a). Using the program CASTp (Dundas, Ouyang et al. 2006), the cleft has a calculated surface area of 1134 $Å^2$ and a volume of 2768 $Å^3$.

A BLAST search reveals that PknH orthologs occur only in pathogenic mycobacteria. Homologous sensor domains in STPKs generally show >50% sequence identity. In addition, the PknH sensor domain is generally <40% identical to mycobacterial LppH proteins, which contain an N-terminal lipid attachment sequence but no kinase domain. Plotting the sequence conservation in PknH orthologs on the sensor-domain structure reveals that residues surrounding the disulfide bonds and lining the bottom of the cleft have a high degree of conservation (Figure 2). In contrast, the residues forming the edges and surface-exposed sides of the cleft are less conserved. These results point to the recent emergence of this fold in mycobacteria and suggest that the sensor-domain homologs are adapted to bind different related ligands in different species.

B. Structural comparisons

A search for similar structures using the DALI server (Holm, Rosenström et al. 2010) indicated that this fold has not been structurally characterized before. The three most related structures found in this search were DIP2269, a hypothetical protein from *Corynebacterium diphtheria* (Z = 9.6, rmsd = 3.6 Å over 118 aligned residues), TM 1622, a GTP-binding regulator from *Thermotoga maritima* (Z = 8.8, rmsd = 3.5 Å over 120 aligned residues) and BT1490, an uncharacterized protein from *Bacterioides thetaiotaomicron* (Z = 8.8, rmsd = 3.2 Å over 119 aligned residues). The modest Z-scores, distinct topologies, high Cα rmsds and uncharacterized activities led to the conclusion that direct structural comparisons between the PknH sensor domain and the DALI search results do not provide insights into the possible functions of this STPK.

Despite this lack of structural homologs, visual comparison between the PknH sensor domain and the glycolipid-binding *Mtb* lipoproteins LprG /Rv1411c (Drage, Tsai et al. 2010; Sieling, Hill et al. 2009) and LppX/Rv2945c (Sulzenbacher, Canaan et al. 2006) shows that all three have a large central cleft bordered on one side by a beta sheet and on the other by two to three alpha helices (Figure 3). The cleft for PknH has a surface area of 1134 $Å^2$ and volume of 2768 $Å^3$, compared to the clefts for LprG (1549 $Å^2$ and 2679 $Å^3$) and LppX (1375 $Å^2$ and 2835 $Å^3$) (Dundas, Ouyang et al. 2006).

LprG is a widely distributed and conserved *Mtb* lipoprotein with TLR2 agonist activity that has been shown to bind the cell wall precursor molecule Ac_1PIM_2 (Drage, Tsai et al. 2010). Phosphatidylmyoinositol mannosides (PIMs) are glycolipids found in the inner and outer membrane of the cell envelopes of all mycobacteria. They consist of a phosphatidylmyoinositol lipid anchor that carries one to six mannose residues with up to four acyl chains (Guerin, Korduláková et al. 2010). In addition to being critical structural components of the cell envelope, PIMs such as Ac_1PIM_2 are precursors for lipomannan and lipoarabinomannan, which modulate host-pathogen interactions over the course of a tuberculosis infection (Mishra, Driessen et al. 2011). LppX is another conserved *Mtb* lipoprotein that shares 31% sequence identity with

LprG and is predicted to bind to phthiocerol dimycocerosates and transport them to the outer layer of the mycobacterial cell envelope (Sulzenbacher, Canaan et al. 2006).

C. PknH function

The PknH STPK sensor domain adopts a novel fold with two intramolecular disulfide bonds and a large v-shaped hydrophobic cleft. Possible functions for the PknH sensor domain can be inferred from its genomic location and reported substrates. The *pknH* gene (Rv1266c) is adjacent to the gene for the EmbR transcriptional regulator (Rv1267) on the *Mtb* chromosome. *In vitro* phosphorylation by PknH enhances EmbR binding the promoter of the *embCAB* arabinosyltransferase genes and leads to increased transcription of these enzymes (Sharma, Gupta et al. 2006). EmbA and EmbB are glycosyltransferases that create the terminal hexaarabinoside motif in *Mtb* cell wall arabinogalactan (Escuyer, Lety et al. 2001) while EmbC synthesizes the arabinan portion of lipoarabinomannan (Zhang, Torrelles et al. 2003). Deletion of PknH in *Mtb* results in a hypervirulent phenotype in mice and decreased transcription of *embB* and *embC* in cultures treated with sublethal concentrations of ethambutol (Papavinasasundaram, Chan et al. 2005). By phosphorylating EmbR, PknH may control the ratio of lipoarabinomannan to lipomannan, a critical determinant of *Mtb* virulence.

In addition, PknH and several other *Mtb* STPKs have been shown to phosphorylate KasA, KasB, and *Mt*FabH. KasA and KasB are β-ketoacyl-ACP synthases that elongate mycolic acid precursors (Molle et al. 2006). Mycolic acids are long-chain (C60-C90) α-alkyl-β-hydroxy fatty acids that promote bacterial resistance to antibiotics and environmental stress (Daffe, Draper et al. 1998) and thus increase *Mtb* virulence (Dubnau, Chan et al. 2000) and persistence (Daffe, Draper et al. 1998). Phosphorylation-induced inhibition of KasA activity presumably leads to immature mycolic acids while phosphorylation-induced stimulation of KasB activity is thought to ensure production of the full-length mycolates required for bacterial survival and virulence (Bhatt, Fujiwara et al. 2007; Bhatt, Molle et al. 2007). *Mt*FabH is the β-ketoacyl-ACP synthase III enzyme that catalyzes the condensation of FAS-I derived acyl-CoAs with malonyl-AcpM, thus linking the FAS-I and FAS-II systems in *Mtb* (Brown, Sridharan et al. 2005). MtFabH is phosphorylated *in vitro* by PknH, PknA, and PknF (Veryron-Churlet, Molle et al. 2009). If PknH functions as a feedback regulator, the sensor domain may be responsive to signals generated in the complex *Mtb* cell wall. Alternatively, PknH may regulate cell-wall production in response to environmental cues, including compounds that are unrelated to the mycobacterial cell wall.

The novel structure of the sensor domain affords few clues about the signaling ligand(s). The putative recognition cleft is narrow, deep, and conserved, as expected for small molecule binding sites in proteins (Liang, Edelsbrunner et al. 1998). Visual comparison between the cleft in PknH and the glycolipid-binding clefts of the *Mtb* lipoproteins LprG and LppX indicates that the PknH binding site is less hydrophobic, making it unlikely that cell-wall glycolipids are the signals. The mixed hydrophobic and polar character of the PknH cleft is consistent with a more polar signal. The stabilizing disulfides impart rigidity to the fold, and the N-terminus, which connects the domain to the transmembrane helix, extends away from the structure. These general characteristics of structural stiffness and loose tethering to the transmembrane helix are shared by the PknD and PknB sensor domains, implying a common signaling mechanism in which ligands may modulate the localization or oligomerization of the kinase.

Table 1: Structural statistics for the x-ray crystal structure of the PknH ECD in complex with Terbium.

Data Collection	**PknH ECD Terbium Complex**
Wavelength (Å)	1.12
Temperature (K)	100
Space group	$P2_1$
Cell parameters	
a *b* *c* (Å)	47.46 35.92 49.31
α β γ (°)	98.36
Resolution (Å) [a]	50.0 - 1.70 (1.76 - 1.70)
R_{sym} (%)	6.1 (25.3)
I/σI	17.4 (4.9)
Completeness (%)	98.8 (98.1)
Redundancy	3.8 (3.8)
SAD Solution	
Protein Copies per a.s.u.	1
Terbium Sites per a.s.u.	2
Figure of Merit	0.424
Refinement	
Resolution (Å)	48.78-1.70
Number of reflections	34190
R_{work}/R_{free} (%)	16.30 / 19.74
Number of atoms	
Protein	1467
Solvent	221
Average B factors	
Protein ($Å^2$)	17
Solvent ($Å^2$)	25
Rmsd	
Bond lengths (Å)	0.013
Bond angles (°)	.960
Ramachandran plot	
Favored (%)	96
Allowed (%)	100
PDB ID	4ESQ

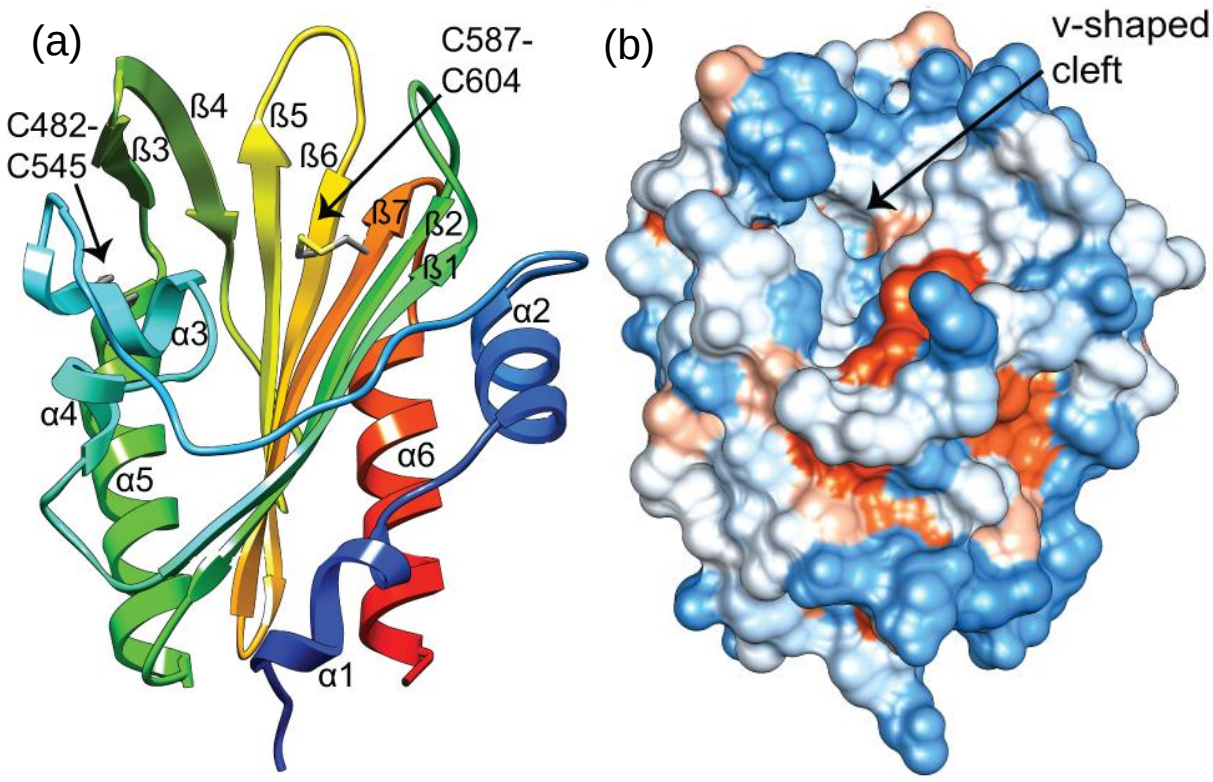


Figure 3.1 Crystal structure of the PknH extracellular sensor domain. (a) Ribbon diagram color-coded from the N-(blue) to the C-terminus (red). **(b)** PknH sensor domain surface colored according to the Kyte and Doolittle hydrophobicity scale (Kyte and Doolittle, 1982), with the most hydrophobic areas in orange and the most polar areas in blue. The large cleft is indicated by a black arrow.

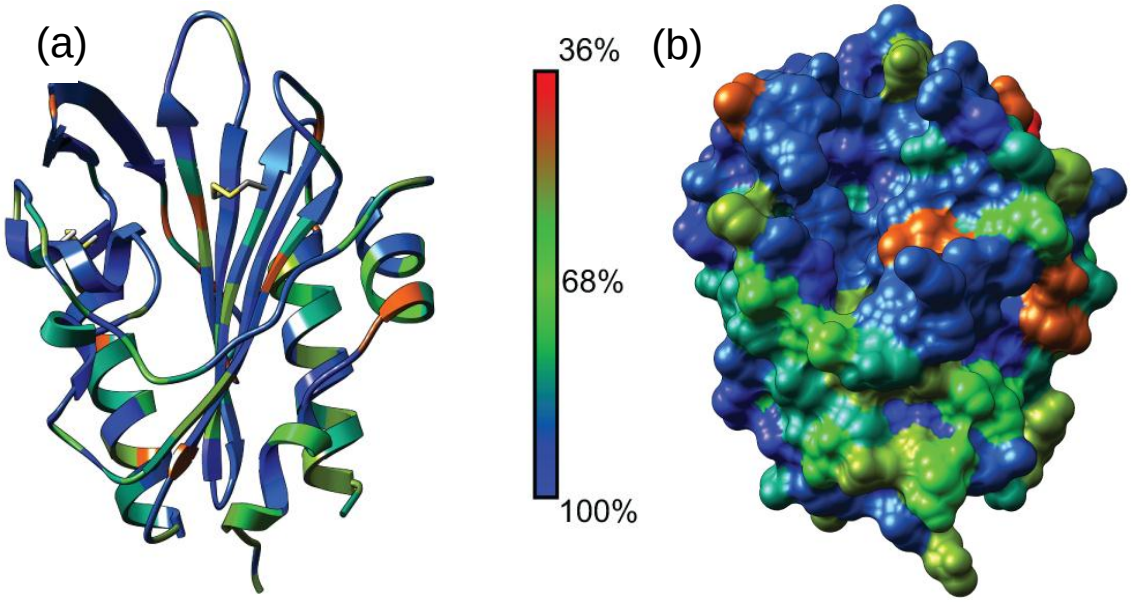


Figure 3.2 Conserved residues in the PknH sensor domain identify the disulfide bonds and the cleft as functional elements. (a) Ribbon representation of the PknH sensor domain with residues colored according to conservation among 11 unique orthologous sequences from different mycobacterial species. The most conserved residues are colored dark blue (100% identity) and the least conserved residues are in red (36% identity). (b) Surface representation of residue conservation.

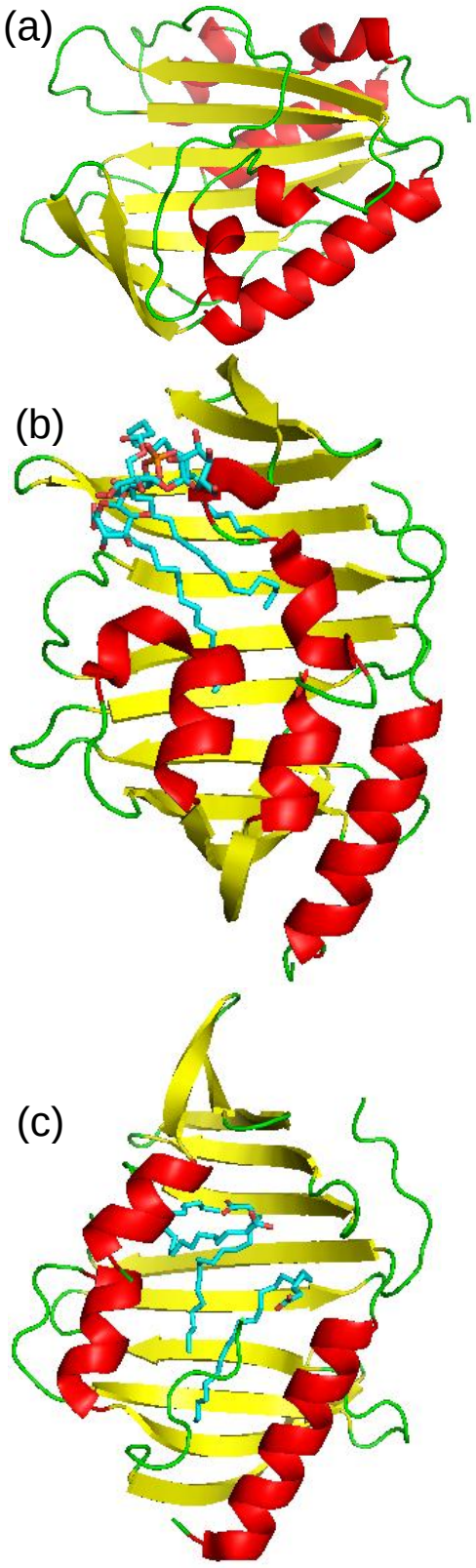


Figure 3.3. **Structures qualitatively resembling the PknH sensor domain.** Ribbon representations of **(a)** PknH turned 90° counter-clockwise from Figure 1, **(b)** *Mtb* lipoprotein LprG bound to the glycolipid cell wall precursor molecule AC_1PIM_2 (PDB 3MHA), and **(c)** *Mtb* lipoprotein LppX bound to two molecules of cis-vaccenic acid (C18:1) and one docosanoic acid (C22:0) molecule (PDB 2BYO).

REFERENCES FOR CHAPTER THREE

Adams, P.D., P.V. Afonine, et al. (2011). The **Phenix** software for automated determination of macromolecular structures. *Methods* **55**(1): 94-106.

Av-Gay, Y. & Everett, M. (2000) The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. *Trends Microbiol* **8**(5): 238-244.

Barthe, P., Mukamolova, G.V. (2010) The structure of PknB extracellular PASTA domain from *Mycobacterium tuberculosis* suggests a ligand-dependent kinase activation. *Structure* **18**(5): 606-615.

Bhatt, A., N. Fujiwara, et al. (2007). Deletion of kasB in *Mycobacterium tuberculosis* causes loss of acid-fastness and subclinical latent tuberculosis in immunocompetent mice. *Proc Natl Acad Sci USA* **104**: 5157-5162.

Bhatt, A., V. Molle, et al. (2007). The *Mycobacterium tuberculosis* FAS-II condensing enzymes: their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. *Mol Microbiol* **64**(6): 1442-1454.

Brown, A.K., S. Sridharan, et al. (2005). Probing the mechanism of the *Mycobacterium tuberculosis* beta-ketoacyl-acyl carrier protein synthase III mtFabH: factors influencing catalysis and substrate specificity. *J Biol Chem* **280**(37): 32539-32547.

Chen, V.B., W.B. Arendall, et al. (2010). **MolProbity**: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 1): 12-21.

Daffe M, and Draper, P. (1998). The envelope layers of mycobacteria with reference to their pathogenicity. *Adv Microb Physiol* **39**: 131-203.

Drage, M.G., H.C. Tsai, et al. (2010). *Mycobacterium tuberculosis* lipoprotein LprG (Rv1411c) binds triacylated glycolipid agonists of Toll-like receptor 2. *Nat Struc Mol Biol* **17**(9): 1088-1095.

Dubnau, E., J. Chan, et al. (2000). Oxygenated mycolic acids are necessary for virulence of *Mycobacterium tuberculosis* in mice. *Mol Microbiol* **36**(3): 630-637.

Dundas, J., Z. Ouyang, et al. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucl Acids Res* **34**(web server issue): W116-W118.

Emsley, P., and Cowtan, K. (2004). COOT: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**(Pt 1): 2126-2132.

Escuyer, V.E., M.A. Lety , et al. (2001). The role of the *embA* and *embB* gene products in the biosynthesis of the terminal hexaarabinosyl motif of *Mycobacterium smegmatis* arabinogalactan. *J Biol Chem* **276**(52): 48854-48862.

Fiuza, M., M.J. Canova, et al. (2008) From the characterization of the four serine/threonine protein kinases (PknA/B/G/L) of *Corynebacterium glutamicum* toward the role of PknA and PknB in cell division. *J Biol Chem* **283**(26): 18099-18112.

Fiuza, M., M.J. Canova et al. (2008). The MurC ligase essential for peptidoglycan biosynthesis is regulated by the serine/threonine protein kinase PknA in *Corynebacterium glutamicum.* *J Biol Chem* **283**(52): 36553-36563.

Galperin, M.Y., R. Higdon, et al. (2010). Interplay of heritage and habitat in the distribution of bacterial signal transduction systems. *Mol Biosyst* **6**(4): 721-728.

Good, M.C., A.E. Greenstein, et al. (2004). Sensor domain of the *Mycobacterium tuberculosis* receptor Ser/Thr Protein Kinase, PknD, forms a highly symmetric β propeller. *J Mol Biol* **339**(2): 459-469.

Guerin, M.E., J. Korduláková, et al. (2010). Molecular basis of phosphatidyl-*myo*-inositol mannoside biosynthesis and regulation in mycobacteria. *J Biol Chem* **285**(44): 33577-33583.

Holm, L. & Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucl Acids Res* **38**(web server issue): W545-W549.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *Mol Biol* **157**(1): 105-132.

Liang, J., H. Edelsbrunner, et al. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Prot Sci* **7**(9): 1884-1897.

Madec, E., A. Laszkiewicz, et al. (2002). Characterization of a membrane-linked Ser/Thr protein kinase in *Bacillus subtilis*, implicated in developmental processes. *Mol Microbiol* **46**(2): 571-586.

Madec, E., A. Stensballe, et al. (2003). Mass spectrometry and site-directed mutagenesis identify several autophosphorylated residues required for the activity of PrkC, a Ser/Thr kinase from *Bacillus subtilis*. *J Mol Biol* **330**(3): 459-472.

Mir, M., J. Asong, et al. (2011). The extracytoplasmic domain of the *Mycobacterium tuberculosis* ser/thr kinase PknB binds specific muropeptides and is required for PknB localization. *PLOS Path.* **7**(7): e1002182 (1-11).

Mishra, A.K., N.N. Driessen, et al. (2011). Lipoarabinomannan and related glycoconjugates: structure, biogenesis and role in *Mycobacterium tuberculosis* physiology and host-pathogen interaction. *FEMS Microbiol Rev* **35**(6): 1126-1157.

Mougous, J.D., C.A. Gifford, et al. (2007). Threonine phosphorylation post-translationally regulates protein secretion in *Pseudomonas aeruginosa*. *Nat Cell Biol* **9**(7): 797-803.

Ortiz-Lombardía, M., F. Pompeo, et al. (2003). Crystal structure of the catalytic domain of the PknB serine/threonine kinase from *Mycobacterium tuberculosis*. *J Biol Chem* **278**(15): 13094-13100.

Otwinoski, Z., and Minor, W. (1997). *Methods Enzymol* **276**: 307-326.

Papavinasasundaram, K.G., B. Chan, et al. (2005). Deletion of the *Mycobacterium tuberculosis pknH* gene confers a higher bacillary load during the chronic phase of infection in BALB/c mice. *J Bacteriol* **187**(16): 5751-5760.

Schnappinger, D., S. Ehrt, et al. (2003). Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. *J Exp Med* **198**(5): 693-704.

Shah, I.M., M.H. Laaberki, et al. (2008). A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. *Cell* **135**(3): 486-496.

Shah, I.M., and Dworkin, J. (2010). Induction and regulation of a secreted peptidoglycan hydrolase by a membrane Ser/Thr kinase that detects muropeptides. *Mol Microbiol* **75**(5): 1232-1243.

Sharma, K., M. Gupta, et al. (2006). Transcriptional control of the mycobacterial *embCAB* operon by PknH through a regulator protein, EmbR, *in vivo*. *J Bacteriol* **188**(8): 2936-2944.

Sieling, P.A., P.J. Hill, et al. (2009). Conserved mycobacterial lipoglycoproteins activate TLR2 but also require glycosylation for MHC class II-restricted T-cell activation. *J Immunol* **180**(9): 5833-5842.

Sulzenbacher, G., S. Canaan, et al. (2006) LppX is a lipoprotein required for the translocation of phthiocerol dimycocerosates to the surface of *Mycobacterium tuberculosis*. *EMBO J* **25**(7): 1436-1444.

Veyron-Churlet, R., V. Molle, et al. (2009). The *Mycobacterium tuberculosis* beta-ketoacyl-acyl carrier protein synthase III activity is inhibited by phosphorylation on a single threonine residue. *J Biol Chem* **284**: 6414-6424.

World Health Organization. (2011). Global tuberculosis control. Accessed online 4/4/2012 at http://www.who.int/tb/publications/global_report/2011/gtbr11_full.pdf.

Young, T.A., B. Delagoutte, et al. (2003). Structure of *Mycobacterium tuberculosis* PknB supports a universal activation mechanism for Ser/Thr protein kinases. *Nat Struct Biol* **10**(3): 168-174.

Zhang, N., J.B. Torrelles, et al. (2003). The Emb proteins of mycobacteria direct arabinosylation of lipoarabinomannan and arabinogalactan via an N-terminal recognition region and a C-terminal synthetic region. *Mol Microbiol* **50**(1): 69-76.

CHAPTER FOUR: SEARCH FOR THE PKNH SIGNALING LIGAND

I. INTRODUCTION

The function of PknH depends critically on the chemical signals received by the sensor domain. Although the sensor domain structure indicates the binding site is more polar than glycolipid binding sites found in lipoproteins such as LprG and LppX (Chapter 3), the identified protein substrates of PknH and its genomic location nonetheless support the hypothesis that PknH regulates cell wall glycolipids or their precursors (Chapter 3). To determine whether PknH binds glycolipids similar to Ac_1PIM_2 bound by the *Mtb* lipoprotein LprG or to PIM bound by LprA (Drage, Tsai et al. 2010), several *in vitro* binding assays were conducted. A set of PknH binding site mutants were created and purified. One of these mutants, PknH V593W, was crystallized, and its structure solved to 1.5 Å resolution. Native gels and carbohydrate stains were used to probe for signs that the PknH sensor domain binds to purified cell wall components. To take a broader approach, the purified PknH sensor domain was incubated with *Mtb* H37Rv lysate followed by mass spectrometry to find potential small molecules that bind PknH. Unfortunately, these essays did not reveal the PknH ligand, and further studies are needed to define the signals that regulate this STPK.

II. MATERIALS AND METHODS

A. Cloning, Expression, and Purification of Proteins

PknH sensor domain residues 435-626, LprG residues 35-236, and LprA residues 26-244 were each amplified from *Mycobacterium tuberculosis* H37Rv genomic DNA and cloned into pET28-based destination vector pHGWA using Gateway enzymes (Invitrogen). The pHGWA vector has an N-terminal His_6 tag followed by a TEV protease cleavage site (Busso, Delagoutte-Busso et al. 2005). The tag leaves Gly-His at the N-terminus of the protein after removal. Mutagenesis was carried out using the QuickChange method (Stratagene). All constructs and mutants were confirmed by DNA sequencing.

The PknH WT and the PknH V593W proteins were expressed, refolded, and purified as described in Chapter 3. N-terminally His_6-tagged LprG and LprA were expressed in BL21 (DE3) Codon Plus cells (Stratagene) using auto-induction (Studier 2005). Cells were grown at 37 °C for six hours and then at 22 °C for 18 hours. Cell paste was resuspended in Nickel A Buffer (25 mM Tris-HCl at pH 8.0, 300 mM NaCl, and 0.5 mM TCEP) plus 30 mM imidazole and protease inhibitors. The cells were lysed via sonication and the lysate centrifuged at 16,000 rpm for 30 minutes. The supernatant was loaded onto a 5 mL HiTrap Fast Flow column (GE Healthcare) equilibrated with Nickel A Buffer. The proteins were eluted with Nickel A Buffer plus 300 mM imidazole, concentrated and further purified by size exclusion chromatography using a HiLoad Superdex 75 column in SEC Buffer (200 mM NaCl, 20 mM Tris-HCl pH 8.0).

B. Crystallization and Structure Determination by Molecular Replacement

PknH V593W was concentrated to 5 mg/mL and crystallized at 18 °C by vapor diffusion in a 1:1.5 ratio of protein to crystallization buffer (0.1 M Bis-Tris (pH 6.0), 0.2 M ammonium acetate, 25% PEG 3350). Additive screening of the initial crystal hit led to improved crystals

with the addition of *N*, *N*-dimethylmethanamine oxide (TMAO). Optimal crystals were grown overnight at 18 °C using 300 nL of 5 mg/mL PknH V593W and 450 nL crystallization buffer plus 100 nL of 100 mM TMAO. Crystals were cryoprotected with mother liquor containing 15% v/v glycerol.

The structure of the wild-type PknH sensor domain (PDB ID 4ESQ) was utilized as a search model for molecular replacement using the program phenix.autoMR (Adams, Afonine et al. 2010). Manual model building and refinement (with hydrogens included) were performed iteratively with Coot (Emsley and Cowtan 2004) and Phenix (Adams, Afonine et al. 2010). The structure was solved to 1.5-Å resolution in the $P2_1$ space group with one protein molecule per asymmetric unit (Table 4.1). 98% of the residues were in favored Ramachandran regions with 100% in the allowed regions. The R_{work} and R_{free} values were 15.07% and 17.49%, respectively (Table 4.1).

C. Native Gels and Carbohydrate Stains

Purified *Mtb* H37Rv arabinogalactan (AG), lipoarabinomannan (LAM), and lipomannan (LM) were obtained from Colorado State University/BEI Resources. To assist future researchers, the protocols used by Colorado State University for obtaining AG, LAM, and LM are detailed below.

The arabinogalactan was produced by hydrolyzing *Mtb* mycolic acid-arabinogalactan-peptidoglycan cell-wall fragments (mAGPs) with KOH in methanol and releasing the arabinogalactan from peptidoglycan by mild acid hydrolysis with H_2SO_4. The soluble arabinogalactan was separated from insoluble peptidoglycan by centrifugation, neutralized, dialyzed against water, and dried for shipping.

The lipoarabinomannan and lipomannan were produced by growing *Mtb* to late-log phase (day 14) in glycerol-alanine-salts (GAS) medium, washing the cells with PBS at pH 7.4 and inactivating them via gamma-irradiation. The cells were delipidated, suspended in PBS buffer with 4 % Triton X-114, and lysed via French Press. The lysate was rocked at 4 °C for 18 hours and the insoluble material was removed by repeated centrifugation at 27,000 x g. The Triton X-114 extract was collected, heated to 37 °C to allow for biphasic partitioning, and centrifuged at 27,000 x g. The detergent layer was collected and macromolecules such as lipoarabinomannan and lipomannan were removed by ethanol precipitation. The ethanol-insoluble material was suspended in PBS and the proteins digested and dialyzed out. The crude carbohydrate mixture was fractionated by size-exclusion chromatography. Buffer contaminants were removed by extensive dialysis using endotoxin free water and buffers, and the final LAM and LM solutions were lyophilized and stored dry at -80 °C for shipping (Chatterjee,, Lowell et al. 1992).

PknH sensor domain (435-626) was concentrated to 200 μM and incubated for four hours at 4 °C with AG, LAM, or LM. As a positive control, *E. coli* peptidoglycan was combined with lysozyme from chicken egg white (Sigma-Aldrich). 2 μL of 10x native gel loading dye (225 mM Tris at pH 6.8 with 50% v/v glycerol and 0.05% w/v Bromophenol Blue) was combined with 18 uL of protein:ligand solution and 10 uL samples were loaded onto precast Novex 12% Tris-Glycine gels. The native gels were run for 6 hours at 60 V with a tris-glycine buffer at pH 9.0. The ProQ Emerald Fluorescent Stain (Invitrogen) for glycoproteins was used to visualize potential glycolipid-binding by PknH.

D. Affinity fractionation of TB lysate

Purified PknH, LprG and LprA were incubated with Ni-NTA resin (MCB Labs) that had been equilibrated with Pulldown Buffer (150 mM NaCl, 20 mM Tris-HCl at pH 8.0). Unbound protein was washed away with Pulldown Buffer plus 20 mM imidazole, leaving each protein bound to the Ni-NTA resin at a concentration of 4 mg/mL as measured by SDS-PAGE.

To prepare the *Mtb* cell lysate, gamma-irradiated *Mtb* cells (strain H37Rv) were obtained from Colorado State University/BEI Resources. Three grams of cell pellet were resuspended in 15 mL of Hank's Buffered Salt Solution (137 mM NaCl, 5.4 mM KCl, 250 μM Na_2HPO_4, 440 μM KH_2PO_4, 1.3 mM $CaCl_2$, 1.0 mM $MgSO_4$, and 4.2 mM $NaHCO_3$) plus protease inhibitors Leupeptin and AEBSF. The solution was incubated with 20 mg of lysozyme from chicken egg white (Sigma-Aldrich) for 1 hour at 25 °C and lysed via sonication on ice. Six mL of *Mtb* cell lysate was incubated with 1 mL of protein-bound Ni-NTA resin for four hours at 4 °C. The resin was separated from unbound material by a low spin centrifugation and washed with 100 mM NaCl and 15 mM PIPES pH 6.5 with stepwise increases in imidazole concentration (0 – 40 mM). Proteins were eluted in this buffer containing 400 mM imidazole and the resulting eluates were divided into two equal sets.

E. Intact Mass, LC-MS and Nanospray-ESI-MS

One set of affinity-chromatography samples was analyzed via SDS-PAGE on 12% Tris-Glycine gels (Novex). Protein bands at 14 kD and 24 kD were excised and their intact masses determined using a Finnigan LTQ-FT (Thermo). LC-MS analysis of the excised SDS-PAGE protein bands was conducted by the Proteomics/Mass Spectrometry Laboratory at UC Berkeley. Excised protein bands were digested with trypsin, and the resulting peptides were extracted from the gel matrix, dried and resuspended in Buffer A (5% acetonitrile/ 0.02% heptaflurobutyric acid). A nano LC column was packed in a 100 μm inner diameter glass capillary with an emitter tip. The column consisted of 10 cm of Polaris c18 5 μm packing material (Varian). The column was loaded by use of a pressure bomb and washed extensively with Buffer A. The column was then directly coupled to an electrospray ionization source mounted on a Thermo-Fisher LTQ XL linear ion trap mass spectrometer. An Agilent 1200 HPLC equipped with a split line so as to deliver a flow rate of 30 nl/min was used for chromatography. Peptides were eluted using a 90 minute gradient from buffer A to 60% Buffer B (80% acetonitrile/0.02% heptaflurobutyric acid). The programs SEQUEST and DTASELECT were used to identify peptides and proteins from the *M. tuberculosis* database (Eng, McCormack et al. 1994; Tabb, McDonald et al. 2002).

In order to search for small-molecule ligands of PknH, the second set of *Mtb* pulldown samples was extracted with methanol and subjected to nano-ESI-MS in negative ion mode. Briefly, 12 μg of protein/eluate mix was denatured in 900 μL methanol for ten minutes followed by vortexing. 10 μL of methanol eluate was loaded onto a nanospray tip for negative-mode electrospray ionization mass spectrometry on a Q-T mass spectrometer (Premier, Waters – Manchester, UK). Mobile phase A was hexane:isopropanol (60:40 v/v) with 0.1% v/v formic acid, 0.05% v/v ammonium hydroxide, and 0.05% v/v triethylamine. Mobile phase B was methanol with 0.1% v/v formic acid, 0.05% v/v ammonium hydroxide, and 0.05% v/v triethylamine. At a flow rate of 0.7 mL/min, the binary gradient started at 5% mobile phase B and linearly increased to 15% B in six minutes and held there for another 10 minutes. Lastly, the gradient was increased to 95% B in eight minutes and held there for six minutes before being

reduced to 5% B in 2 minutes. These nano-ESI experiments were processed using Masslynx software.

III. RESULTS AND DISCUSSION

A. PknH V593W Structure

To characterize the binding of PknH to potential ligands, a set of single substitution mutations in the PknH sensor domain was created. Based on analysis of the wild-type PknH sensor domain structure, thirteen sites for useful mutations were chosen (Figure 4.1). Each mutation was chosen because it would presumably alter the cleft surface and thus prevent ligand binding while still permitting the protein to fold. Nine of the 13 variants were successfully cloned, expressed, and refolded according to the protocol described in Chapter 3 for the wild-type PknH sensor domain (data not shown). The most highly expressed and well-behaved PknH mutant was V593W, which crystallized under the same conditions as the wild-type PknH sensor domain. A native X-ray data set was collected, and the structure was solved to 1.5-Å resolution using the initial PknH WT structure as a template for molecular replacement (Figure 4.2 and Table 4.1). Unfortunately, V593 does not face the cleft and the Trp side chain is accommodated on the surface (Figure 4.2 (B)).

B. Native PAGE and Glycoprotein Staining

To probe for PknH binding to purified *Mtb* cell wall components such as lipomannan or lipoarabinomannan, I incubated purified PknH sensor domain its V593W mutant in a 1:1 ratio with these components and analyzed the samples via native PAGE for signs of binding (Figure 4.3). A comparison of Lanes 4, 5, and 8 of the native gel in Figure 4.3(A) indicates the migration and appearance of the PknH protein band was not shifted when purified lipomannan or lipoarabinomannan were added (Figure 4.3(A)). Similar native PAGE results were obtained for purified *Mtb* arabinogalactan, PIM_2, and PIM_6 (data not shown). A doubling of the amount of purified cell wall component had no observable effect on similar native gel experiments (data not shown). In contrast, when lysozyme was incubated with *E. coli* peptidoglycan (Figure 4.3(A), lane 10), there is a smeared but visible protein band indicating lysozyme is binding peptidoglycan and pulling it into the gel matrix (Figure 4.3 (A)).

Staining with the ProQ Emerald Glycoprotein stain involves the use of periodic acid to turn glycols into aldehydes, followed by incubation of the protein with a proprietary fluorescent conjugate the gives off a visible green color when illuminated at 300 nm (Hart *et al*., 2003). After the native gel in Figure 4.3(A) was stained for glycoproteins, a faint fluorescent signal for the PknH ECD is visible in the stained gel (Figure 4.3(B), lane 4). Although glycosylation does occur in bacteria (Nothaft *et al.*, 2010) this faint staining is likely due to non-specific binding of the ProQ dye to PknH. This stain confirmed that lipomannan or lipoarabinomannan did not change the migration of the PknH ECD on a native gel (Figure 4.3 (B), lanes 4, 5, and 8), while the positive control peptidoglycan/lysozyme lane indicates this protein did bind its substrate and move it into the gel (Figure 4.3(B), lane 10). Thus, lipomannan and lipoarabinomannan do not appear to be PknH ligands. The lack of PknH ECD binding to these purified cell wall components could be due to the fact that the native gel buffer was set at pH 9.0; it is possible that binding between PknH and its ligand only occurs at a pH equal to what the bacterium encounters

inside activated immune cells (pH 6.76 – 7.50; Vandal *et al*., 2008) or at a lower pH values that occur in acidic phagosomes. Alternatively, the purified components may need to be processed into smaller fragments by bacterial or host enzymes prior to binding the PknH sensor domain. As a final possibility, the ligand for PknH may be a molecule produced solely by host cells or a specific subset of host cells after infection.

C. Pulldowns Using His_6-tagged Proteins with *Mtb* H37Rv Lysate

To increase the likelihood of obtaining a properly sized or properly processed cell wall component that would bind to the PknH sensor domain, affinity-chromatography experiments using purified His_6-tagged PknH, V593W, and the *Mtb* lipoproteins LprG and LprA were conducted using *Mtb* H37Rv cell lysate. After four hours of incubating the purified proteins with the *Mtb* lysate, the resin with the bait proteins was washed with increasing amounts of imidazole (0-40 mM) to reduce non-specific binding of contaminants. Analysis of the third wash fractions (Figure 4.4 (A), all lanes marked "W3') showed a protein band with an apparent size of 12-14 kD appears in every sample, including the beads/negative control (Figure 4.4 (A)). By the fifth wash step, this contaminant was not visible in the LprG and LprA samples, but was still visible in the PknH WT and V593W samples (Figure 4.4 (A)). After seven washes, 300 mM imidazole was used to elute the protein samples from the Ni-NTA resin and the eluates were analyzed by SDS-PAGE and nano-ESI-MS in negative mode. Aside from the expected bands for the bait proteins at 24-26 kDa (Figure 4.4 (B)), a band of 12-14 kD is visible in the SDS-PAGE gel in the PknH WT and V593W elutions (Figure 4.4 (B), lanes "PknH/E" and "V593W/E").

Both of these protein bands were excised and subjected to intact mass spectrometry. The larger-sized band was made up of three species with masses of 24064, 24242, and 24455 Da (Figure 4.5). The species with a mass of 24064 Da was much more abundant in the sample than the other species (Figure 4.5). The expected mass for the N-terminally His_6-tagged PknH WT protein is 24199.9 Da. The less abundant peaks at 24242 and 24455 Da might represent modifications to the PknH sensor domain that occurred during its expression in *E. coli* or its incubation with the *Mtb* lysate. Assuming the full-length PknH sensor domain was intact, then the modification/s added a mass of 255.1 Da (24455 – 24199.9 = 255.1), or 42.1 Da (24242 – 24199.9 = 42.1). A search of the literature has not turned up any modifications that add 255.1 daltons or 42.1 Da to proteins.

The PknH eluates were further analyzed via LC-MS to try and pinpoint the identity of the 14 kDa contaminant. The list of candidate co-eluting proteins is found in Table 2. There was no perfect match between the proteins identified by LC-MS and the 14035 Da contaminant, but the most likely candidate is HspX (Rv2031c) (Table 2), a 16-kDa α-crystallin-like protein and immunodominant antigen (Lee, Hefta et al. 1992) that has been shown to act as a molecular chaperone (Yang, Huang et al. 1996).

Aside from the 14-kD contaminant, no bands for the proteins identified in Table 2 are visible in Figure 4.4, indicating the other proteins identified via LC-MS as potential PknH interactors did not bind strongly to PknH and the observed interactions may not be biologically relevant. This conclusion is also consistent with the structural data presented in Chapter 3, which suggests that the sensor domain binds a small molecule ligand. As expected, no protein bands were visible in the elution samples for LprG (Figure 4.4(B)), which specifically binds

Ac_1PIM_2 and closely related glycolipids (Drage, Tsai et al. 2010), and for LprA (Figure 4.4(B), lane 9), which has 31% sequence identity to LprG and is also thought to bind similar classes of glycolipids (Drage, Tsai et al. 2010).

To identify potential small molecule ligands for PknH, the *Mtb* pulldown elution fractions for PknH and LprG were extracted with methanol and analyzed via nanospray ESI-MS (Figure 4.6(A) and (B)). In accord with a similar set of experiments performed on LprG and LprA by Drage and coworkers (Drage, Tsai et al. 2010), ions corresponding to the predicted masses for phosphoinositol (m/z 851.6) and Ac_1PIM_2 (m/z 1413.1) were identified in the LprG elution sample (Figure 4.6(B)). In contrast, neither of these ions was present in the PknH sample (Figure 4.6(A)), which had a different ion fragmentation pattern. Based on comparisons between the mass chromatograms for PknH, LprA and the beads-only negative control samples (data not shown), it was determined that the peaks present in the PknH sample are contaminants and do not represent potential ligands. This lack of glycolipid-based signal co-eluting with PknH is in line with our structural evidence that the PknH ECD cleft is more polar than the cleft in LprG (Chapter 3).

IV. CONCLUSION

To further understand the mechanisms that regulate STPKs and cell wall architecture in *Mtb*, I biochemically and structurally characterized the extracellular sensor domain of the receptor STPK, PknH. Initial analysis of recombinantly produced PknH sensor domain by size-exclusion chromatography and small-angle x-ray scattering indicated the protein was soluble but completely unfolded. On-column refolding using decreasing amounts of guanidinium hydrochloride led to properly refolded protein that crystallized as a monomer with a completely novel protein fold made up of six alpha helices and seven beta strands. The PknH sensor domain had a large and deep cleft with a mixed polar and hydrophobic nature and two disulfide bonds. These results indicated PknH binds a small molecule ligand which may affect its quaternary structure and/or its localization.

In order to determine what ligand binds to the PknH sensor domain, I conducted native gel binding assays using purified *Mtb* arabinogalactan, lipoarabinomannan, and lipomannan. These assays indicated that *in vitro*, the PknH sensor domain does not bind to any of these purified cell wall components. Affinity chromatography of His_6-tagged PknH incubated with *Mtb* H37Rv cell lysates followed by intact mass, LC-MS, and ESI-MS analysis indicated that PknH does not bind glycolipids such as phosphoinositol or Ac_1PIM_2. Although the ligand for PknH was not identified, these studies have set the stage for future researchers to identify the ligand and discover the function of PknH.

Table 4.1. X-ray data collection, analysis and refinement statistics for the structure of PknH V593W.

Data Collection	**PknH V593W**
Wavelength (Å)	1.12
Temperature (K)	100
Space group	$P2_1$
Cell parameters	
a b c (Å)	44.66, 35.22, 48.36
α γ β (°)	90 93.76 90
Resolution (Å) [a]	50.0 - 1.50 (1.54 - 1.50)
R_{sym} (%)	8.6 (30.6)
I/σI	18.1 (2.8)
Completeness (%)	97.2 (76.3)
Redundancy	5.5 (2.8)
Refinement	
Protein Copies per a.s.u.	1
Resolution (Å)	48.25-1.50
Number of reflections	23661
R_{work}/R_{free} (%)	15.07 / 17.49
Number of atoms	
Protein	3063
Solvent	210
Average B factors	
Protein ($Å^2$)	16
Solvent ($Å^2$)	25
Rmsd	
Bond lengths (Å)	0.006
Bond angles (°)	1.130
Ramachandran plot	
Favored (%)	98
Allowed (%)	2

Note to Table 4.1: Data were collected at 100 K using Beamline 8.3.1 at the Lawrence Berkeley National Laboratory Advanced Light Source (MacDowell, Celestre et al. 2004). Data were reduced and scaled with HKL2000 (Otwinowski and Minor, 1997). The structure was determined using Phenix.automr (Adams, Afonine 2010) and the model was adjusted manually using Coot (Emsley and Cowtan 2010). Building and refinement (which included hydrogens) were completed with Phenix.refine and Coot.

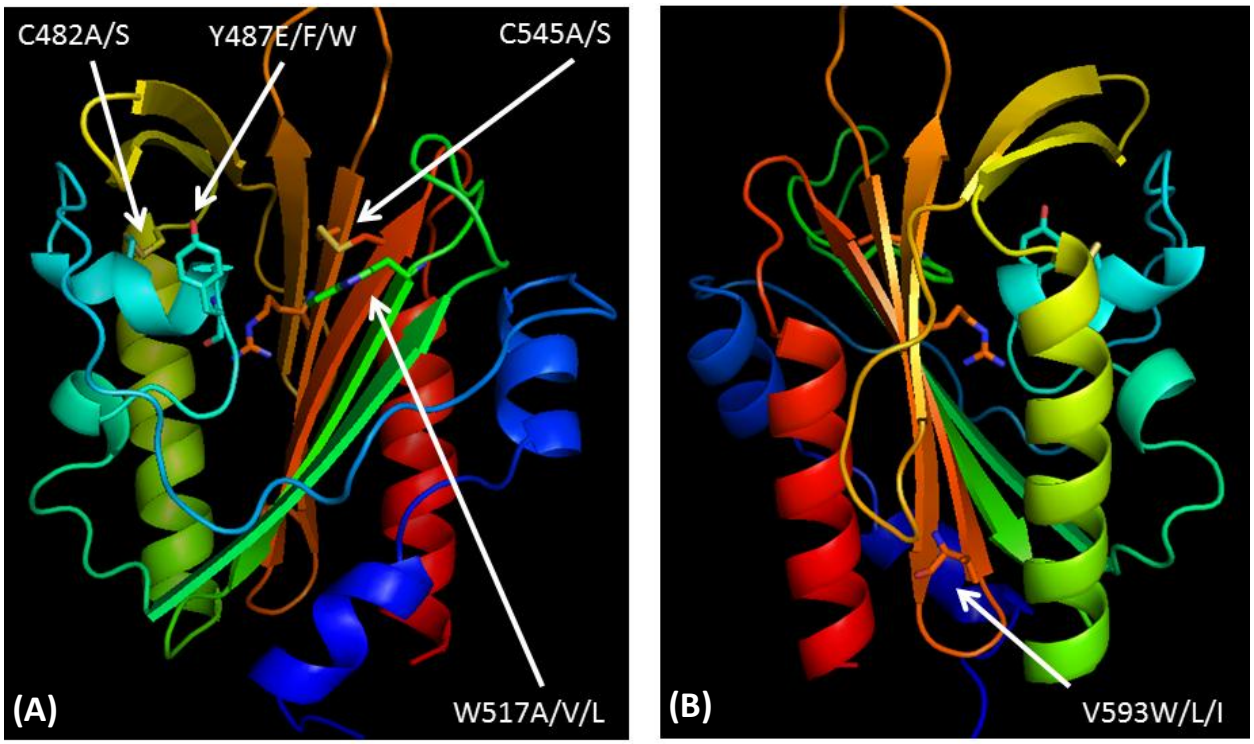


Figure 4.1 (A) Ribbon representation of the PknH sensor domain structure with mutations indicated as sticks. **(B)** Rear view of (A) highlighting V593.

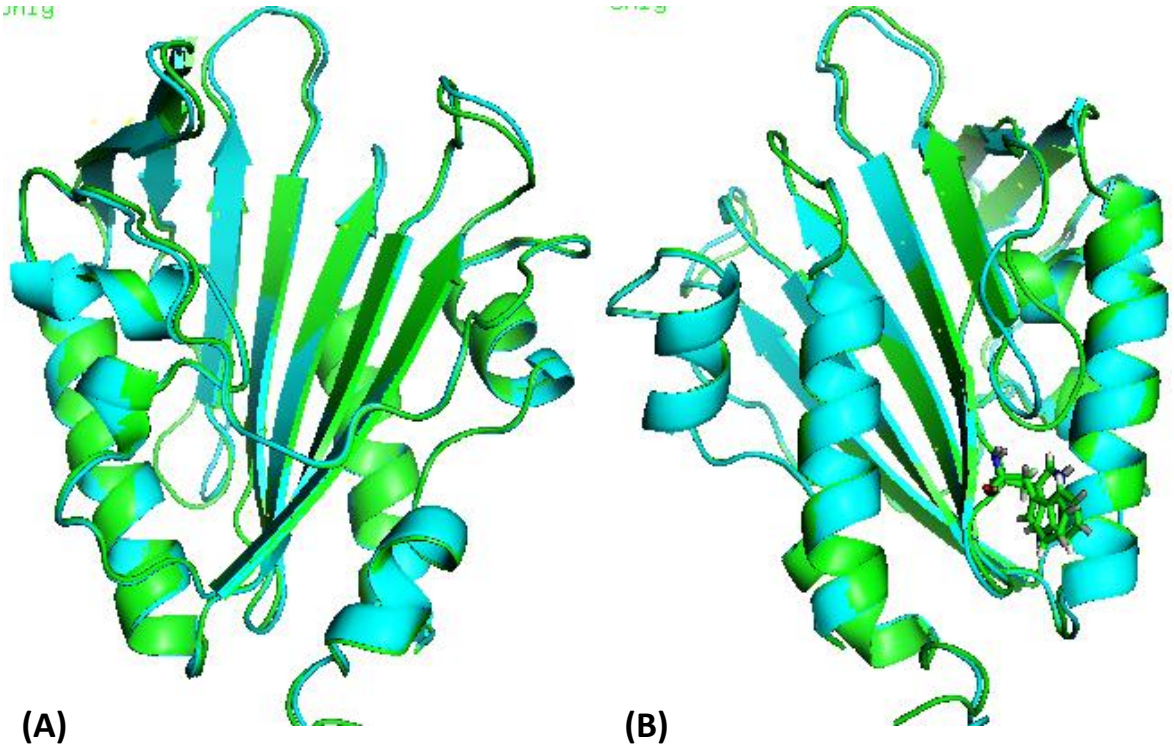


Figure 4.2 (A) Ribbon representation of the wild-type PknH sensor domain structure (light blue) superimposed on the structure of the V593W mutant (green). **(B)** Rear view of **(A)** with the V593W substitution shown as sticks.

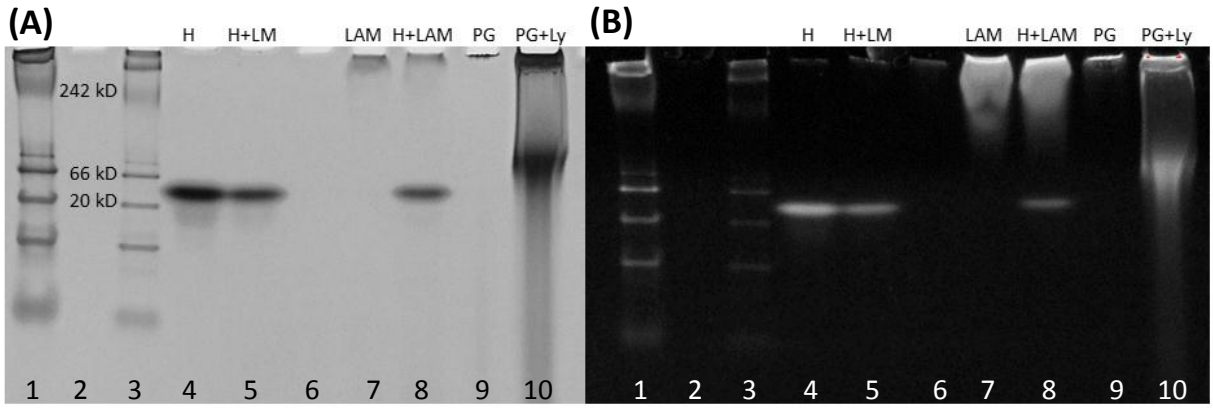


Figure 4.3 (A) A native gel with PknH sensor domain and purified *Mtb* cell wall components. Lanes 1 and 3 contain 8 and 4 uL of NativeMark protein ladder (Invitrogen). Lane 4 = PknH sensor domain. Lane 5 = PknH plus lipomannan (LM) in a 1:1 ratio of protein: LM. Lane 6 = blank. Lane 7 = Lipoarabinomannan (LAM) in isolation. Lane 8 = PknH plus LAM (1:1 ratio). Lane 9 = *E. coli* peptidoglycan in isolation. Lane 10 = *E. coli* peptidoglycan plus lysozyme as a positive control for cell wall component binding. **(B)** Native gel from (A) after carbohydrate staining with Invitrogen ProQ Emerald Glycoprotein stain.

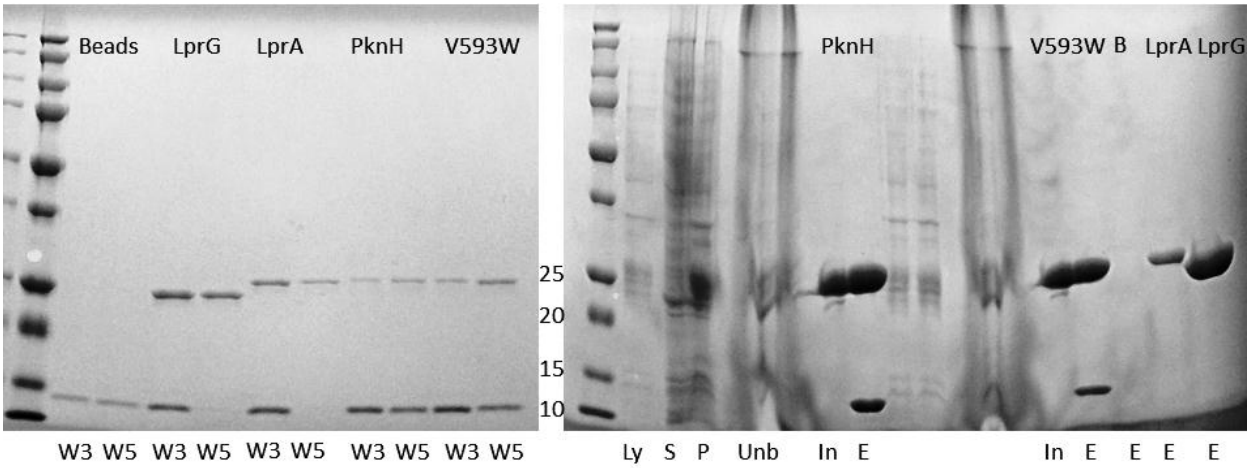


Figure 4.4. Affinity chromatography of *Mtb lysate* on the PknH sensor domain

(A) SDS-PAGE of the wash fractions using *Mtb* H37Rv lysate. In this gel, "W3"and "W5" = the third and fifth wash fractions prior to elution of bound proteins from the Ni-NTA resin. "Beads" = Ni-NTA beads used as a negative control for binding. "LprG" = positive control for glycolipid binding. "LprA" = Rv1270c, another *Mtb* lipoprotein thought to bind glyclolipids in a similar fashion to LprG. "PknH" = PknH sensor domain . "V593W" = PknH sensor domain mutant.

(B) SDS-PAGE demonstrating the refolding and purification of the PknH sensor domain protein used for the affinity chromatography. "L" = the lysate from *E. coli* cells used to produce PknH WT sensor domain. "S" = supernatant from *E. coli* cells. "P" = pellet from *E. coli* cells. "Unb" = unbound fraction during PknH sensor domain refolding/purification with 6M guanidine hydrochloride. "PknH/ln" = PknH sensor domain after refolding and purification, but prior to incubation with *Mtb* lysate. "PknH/E" = elution fraction for PknH sensor domain after incubation with *Mtb* H37Rv lysate. "V593W" = PknH sensor domain with V593W mutation after refolding and purification, but prior to incubation with *Mtb* lysate. "V593W/E" = elution fraction for V593W mutant after incubation with *Mtb* H37Rv lysate. "B" = blank lane. "LprA/E and LprG/E" = Elution fractions after incubation of LprG and LprA with *Mtb* H37Rv lysate.

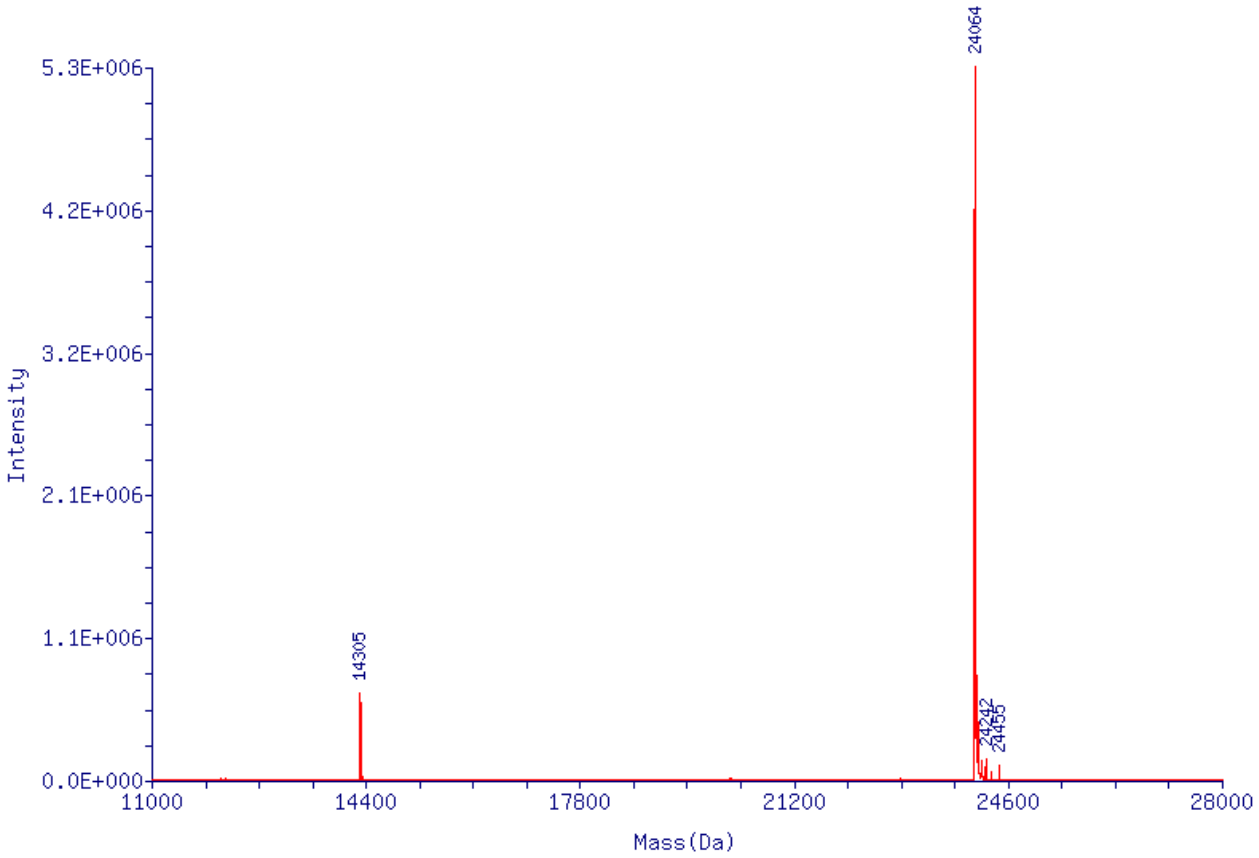


Figure 4.5 Intact mass spectrometry results obtained after cutting out two bands from Figure 4.4(B), lane "PknH/E". The smaller protein was found to have a mass of 14035 Da, while the larger protein band, presumably His6-tagged PknH ECD, was made up of three different species with masses at 24064, 24242, and 24455 Da.

Mtb Rv Gene Number	Sequence Count	Spectrum Count	Sequence Coverage %	Protein Length (aa)	Molecular Mass (kD)	Brief Description Of Function
1738	2	4	16.0	94	10.606	Hypothetical protein
0705	2	2	8.6	93	10.804	Ribosomal protein
2442c	4	8	29.8	104	11.151	50s ribosomal protein
0636	6	11	46.5	142	14.934	Hypothetical protein
2031c	4	8	26.4	144	16.227	HspX/14 kD antigen
3688c	2	2	13.0	154	16.701	Hypothetical protein
1388	2	2	12.6	190	20.835	Integration host factor
0671	2	3	2.5	280	29.579	LpqP lipoprotein
2678c	3	4	2.5	357	37.606	Decarboxylase
0896	2	4	6.3	431	47.978	Citrate synthase
1266c	14	20	12.6	626	66.754	PknH
0101	2	2	1.3	2512	269.406	Peptide synthetase

Table 4.2. *Mtb* H37Rv proteins identified via LC-MS after affinity chromatography using refolded His_6-tagged PknH sensor domain (residues 435-626) incubated with *Mtb* H37Rv lysate. The proteins are sorted by predicted molecular mass (kD) for the full-length protein.

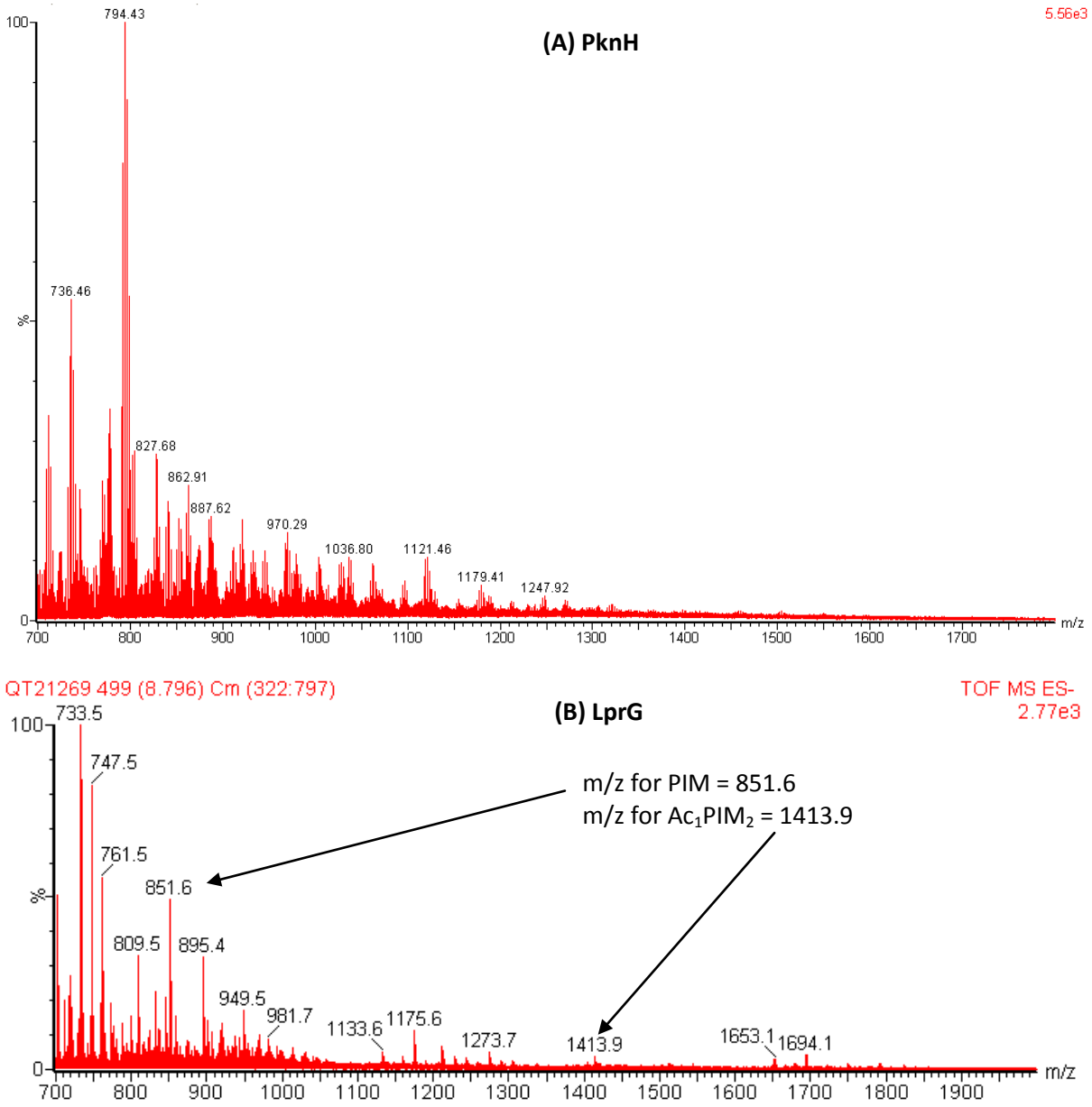


Figure 4.6. Negative-mode ESI-MS results for (A) the PknH sensor domain and (B) LprG/Rv1411c. The predicted m/z values for the glycolipid PIM (851.6) and the glycolipid Ac_1PIM_2 (1413.9) are indicated, with arrows pointing to their corresponding peaks in the LprG sample.

REFERENCES FOR CHAPTER FOUR

Adams, P.D., P.V. Afonine, et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 2): 213-231.

Busso, D., B. Delagoutte-Busso, et al. (2005). Construction of a set of Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*. *Anal Biochem* **343**(2): 313-321.

Chatterjee, D., K. Lowell, et al. (1992). Lipoarabinomannan of *Mycobacterium tuberculosis*. Capping with mannosyl residues in some strains. *J Biol Chem* **267**(9): 6234-6239.

Emsley, P., and Cowtan, K. (2004). COOT: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**(Pt 1): 2126-2132.

Eng, J.K., A.L. McCormack, et al. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**(11): 976-989.

Lee, B.Y., S.A. Hefta, et al. (1992). Characterization of the major membrane protein of virulent *Mycobacterium tuberculosis*. *Infect Immun* **60**(5): 2066-2074.

MacDowell, A.A., R.S. Celestre, et al. (2004). Suite of three protein crystallography beamlines with single superconducting bend magnet as the source. *J Synchrotron Radiat* **11**(Pt 6): 447-455.

Nothaft, H. and C.M. Szymanski. (2010). Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* **8**(11): 765-778.

Tabb, D.L., W.H. McDonald, et al. (2002). DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **1**(1): 21-26.

Yang, H., S. Huang, et al. (1999): The *Mycobacterium tuberculosis* small heat shock protein Hsp 16.3 exposes hydrophobic surfaces at mild conditions: conformational flexibility and molecular chaperone activity. *Protein Sci* **8**(1): 174-179.

APPENDIX A

INTERDOMAIN COMMUNICATION IN THE *MYCOBACTERIUM TUBERCULOSIS* ENVIRONMENTAL PHOSPHATASE Rv1364c

Andrew E. Greenstein[1,2], Michal Hammel[3], Alexandra Cavazos[1], and Tom Alber[1]

[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720-3220, USA

[2]Current address: Gilead Sciences, 333 Lakeside Drive, Foster City, CA 94404, USA

[3]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Running Title: The *Mtb* phosphatase, Rv1364c

Key words: PP2C phosphatase, SAXS, partner switching, transcriptional regulation, Rv1364c

Address correspondence to: Tom Alber, University of California, Berkeley, 374B Stanley Hall # 3220, Berkeley, CA 94720-3220. (510) 642-8758; email: tom@ucxray.berkeley.edu

An "environmental phosphatase" controls bacterial transcriptional responses through alternative sigma factor subunits of RNA polymerase, and a partner switching mechanism has been proposed to mediate phosphatase regulation. In many bacteria, the environmental phosphatase and multiple regulators are encoded in separate genes whose products form transient complexes. In contrast, in the *Mycobacterium tuberculosis* homolog, Rv1364c, the phosphatase is fused to two characteristic regulatory modules with sequence similarities to anti-sigma factor kinases and anti-anti-sigma-factor proteins. Here we exploit this fusion to explore interactions between the phosphatase and the regulatory domains. We show quantitatively that the anti-sigma factor kinase domain activates the phosphatase domain, the kinase-phosphatase fusion protein autophosphorylates in *E. coli*, and phosphorylation is antagonized by the phosphatase activity. Small angle x-ray scattering defines solution structures consistent with the interdomain communication observed biochemically. Taken together, these data indicate that Rv1364c provides a single-chain framework to understand the structure, function and regulation of environmental phosphatases throughout the bacterial kingdom.

Introduction

One third of the world's population is seropositive for *Mycobacterium tuberculosis* (*Mtb*). While most anti-microbial therapeutics target actively growing bacteria, *Mtb* can evade current drugs by converting to a persistent, metabolically suppressed state (1). Moreover, *Mtb* survives in distinct environments *in vivo* by adjusting transcriptional programs. Little is known about the transcriptional cues that mediate such developmental transitions, and defining their underlying mechanisms remains the focus of much mycobacterial research (2-6). The *Mtb* genome encodes 13 sigma-factor homologs, several of which play essential roles in disease (7).

Diverse bacteria employ sigma factors to mediate the transcriptional programs that drive life-cycle transitions (8,9). Alternative sigma factors, unique RNA-polymerase subunits that mediate promoter choice, are known to respond to environmental cues such as heat and limited energy (10) to drive the transition into spore formation (11) and stationary phase (12). As reviewed in (9), anti-sigma factors exemplified by the *Bacillus subtilis* Regulator of Sigma B W (RsbW) are capable of binding sigma factors and sequestering them away from RNA polymerase (Fig. 1). Through a partner switching mechanism, anti-anti sigma factor proteins such as RsbV bind the anti-sigma factors, freeing the cognate sigma factors to activate transcription. Anti-sigma factors, which display homology to histidine kinases, phosphorylate their anti-anti-sigma-factor antagonists on a serine or threonine residue (9). A master "environmental phosphatase", such as RsbU in *B. subtilis*, reverses this phosphorylation and restores the complexes of anti- and anti-anti-sigma factors (13).

These RsbU-like Ser/Thr phosphatases are regulated by proteins with remarkable homologies to the anti- and anti-anti sigma factors in a manner that shows parallels to and differences from sigma-factor regulation (Fig. 1; (13)). An upstream anti-sigma factor homolog, such as RsbT, binds the phosphatase and, in contrast to the downstream sigma-factor antagonism, stimulates the phosphatase activity. Proteins with homology

to anti-anti sigma factors, like RsbS, bind the upstream anti-sigma factor and prevent phosphatase stimulation. RsbS does not bind RsbT in isolation, but rather forms a large complex with multiple upstream anti-anti-sigma factors, such as RsbRA, RsbRB, RsbRC, and RsbRD in the so-called "stressosome" (14), to sequester RsbT. Phosphorylation of RsbR and RsbS by the cognate anti-sigma factor kinase releases RsbT. RsbS is dephosphorylated by the Ser/Thr phosphatase RsbX, which displays homology to RsbU and the PP2C phosphatases (15).

The high degree of sequence similarity between downstream and upstream anti- and anti-anti sigma factors despite their distinct functions has presented challenges for deciphering the mechanisms of anti-anti- and anti-sigma factor functions. Moreover, biochemical analysis has been impeded because most of these regulators, when expressed in isolation, are insoluble. This property reflects the functions of these regulators in multi-protein complexes. Co-expression studies are difficult to design because binding partners cannot be predicted reliably based on homology.

In the RsbU-like environmental phosphatase of *M. tuberculosis* (Rv1364c), however, the catalytic domain is fused in a single polypeptide chain to regulatory modules. A PAS domain occurs at the N-terminus, and an anti-sigma-factor and an anti-anti-sigma-factor domain follow the phosphatase module. Such a combination of domains within a single polypeptide has not been described outside of Rv1364c and close mycobacterial orthologs. We hypothesize that the anti- and anti-anti-sigma factor domains of Rv1364c function like the *B. subtilis* RsbT/RsbS pair to regulate the phosphatase domain (16). These intramolecular interactions in Rv1364c may be representative of the transient and biochemically intractable multi-protein complexes found in other bacteria (17). Thus, Rv1364c may serve as a model system of the environmental phosphatase regulatory circuit ultimately responsible for sigma factor regulation.

To explore the mechanisms of environmental phosphatase regulation, we probed the function of Rv1364c variants biochemically. We found that Rv1364c encodes an active phosphatase that is stimulated by the anti-sigma factor domain. Moreover, contrary to a recent report (18), the anti-sigma factor domain functions as an active kinase. Mutational inactivation of the phosphatase catalytic site stabilized phosphorylated Rv1364c, while mutations of the kinase active site or the predicted phospho-acceptor residue blocked phosphorylation. To experimentally characterize the structural changes driven by phosphorylation, we employed small-angle X-ray scattering (SAXS) to define the shapes and conformations of the unphosphorylated and phosphorylated forms of Rv1364c. Rv1364c typifies functionally flexible systems that are suitable for SAXS analyses (19,20). The solution structural envelopes suggest that Rv1364c forms an elongated dimer. Stable phosphorylation resulting from a mutation in the phosphatase active site caused a structural change consistent with the rearrangement of domains within the dimer. This conformational change provides a molecular context for the interdomain communication and phosphoregulation implied by the biochemical studies.

Experimental Procedures

DNA constructs--Start and end points for each domain of the Rv1364c gene were determined by alignment with homologs identified using PHYRE (21) and BLAST. Using *Mtb* H37Rv genomic DNA as a template for PCR amplification, gene segments encoding the Rv1364c gene were inserted into the Gateway vector pHMGWA (22), which included NH_2-terminal 6X-His and maltose binding protein (MBP) tags, followed by a tobacco etch virus (TEV) protease site. Mutations were introduced using the QuikChange method (Stratagene). All constructs were confirmed by DNA sequencing.

Protein expression and purification—BL21(DE3) Codon Plus cells (Stratagene) cells harboring each expression plasmid were grown at 37 °C to an absorbance (A_{600}) of 1.8 in Terrific Broth (Research Products International), moved to 18 °C for 10 min, induced using 300 μM isopropyl 1-thio-β-D-galactopyranoside, and grown for an additional 4 h. The cells were harvested by centrifugation and resuspended in 100 mM Tris-HCl (pH 7.8), 0.3 M NaCl, 10% (v/v) glycerol, and 0.5 mM Tris(2-carboxyethyl)phosphine

hydrochloride (TCEP). Cells were lysed by sonication on ice, and the lysate was clarified by centrifugation. The supernatant was loaded onto a 5-ml chelating Sepharose HP column (Amersham Biosciences) equilibrated with 0.1 M $NiSO_4$ and eluted in lysis buffer with 300 mM imidazole. Protein was dialyzed into 50 mM HEPES (pH 7.65), 150 mM NaCl, 0.5 mM TCEP, and 5 mM $MnCl_2$ and simultaneously proteolyzed with TEV protease overnight at 4 °C. Each protein was further purified by gel-filtration chromatography using a HiLoad 26/60 Superdex 75 column (Amersham Biosciences) or Sephacryl S-200 column (Amersham Biosciences). Elution profiles were compared to gel filtration protein standards (BioRad) for apparent molecular weight calculations. Protein was dialyzed into 100 mM NaCl, 15 mM HEPES 7.6, 0.5 mM TCEP, 2 mM $MnCl_2$, and 10% glycerol prior to DLS and SAXS analysis.

Phosphatase activity assays—Kinetic assays were performed with 860 nM enzyme diluted into 100 mM Tris pH 8, 50 mM NaCl, 0.5 mM TCEP, 10% glycerol, 10 mM $MgCl_2$, and 10 mM $MnCl_2$ with 0.78-50 mM p-nitrophenylphosphate (pNPP) in a total volume of 100 μL. pNPP hydrolysis kinetics were monitored continuously for 60 min at 405 nM in triplicate. Curve fitting and kinetic-constant calculations were performed using SigmaPlot (Systat Software, Inc). For pH-dependence end point assays, 860 nM full-length Rv1364c was incubated in 100 mM NaCl, 10% glycerol, 10 mM $MgCl_2$, 10 mM $MnCl_2$, 25 mM pNPP and 50 mM buffer. The following buffers were used: sodium cacodylate, pH 5.6, 6.0, and 6.4; HEPES pH 6.8, 7.2, and 7.6; and Tris pH 8.0. Reactions were allowed to proceed for 10, 20, 30, or 40 minutes before quenching with the addition of 25 μL 500 mM ethylenediaminetetraacetic acid (EDTA, pH 8.0) and 25 μL 1 N sodium hydroxide.

Phosphorylation state detection—Phosphorylation reactions were performed in 15 mM HEPES pH 7.6, 0.1 M NaCl, 2 mM $MgCl_2$, 2 mM $MnCl_2$, 2 mM ATP, 10% glycerol for 90 min. Each Rv1364c construct was desalted prior to electrophoresis on a Tris-glycine gel (Invitrogen) and stained in parallel with either Coomassie blue or the Diamond Pro-Q stain (Invitrogen). Fluorescence was quantified on a Typhoon 8600 with ImageQuant (GE Healthcare). Serial dilutions of the phopshorylated *Mtb* PknB kinase domain were used to ensure phosphorylation was detected in the linear range.

SAXS data collection and evaluation-- SAXS data were collected at the SIBYLS beamline at the Lawrence Berkeley National Laboratory Advanced Light Source (19). Using a wavelength, λ, of 1.03 Å with the sample-to-detector distance set to 1.5 m resulted in scattering vectors, q, ranging from 0.01 $Å^{-1}$ to 0.30 $Å^{-1}$. The data were acquired at 20 °C, and short and long time exposures (0.5 s, 5 s) were merged for the calculations using the entire scattering profile. Data were processed as described (19). The experimental SAXS data were measured at different protein concentrations to explore multimer formation and aggregation using Guinier plots. The radius of gyration R_G was derived by the Guinier approximation $I(q)=I(0)\ \exp(-q^2R_G^2/3)$ with the limits $qR_G<1.3$. The interference-free SAXS profile was estimated by extrapolating the measured scattering curves to infinite dilution.

The SAXS curves measured for different concentrations (1-4 mg/ml) of wild-type and Asp328Ala Rv1364c displayed a concentration-dependence arising from self-association in the concentrated samples. The estimated infinite dilution SAXS represents a form factor of the interference-free state. The GNOM programs (23) were used to compute the pair-distance distribution functions, P(r), and the maximum dimension of the macromolecule, D_{max}. The overall shapes were restored from the experimental data using the program GASBOR (24). Sixteen low-resolution models obtained from different runs were averaged using the program DAMAVER (25) to construct the average model representing the general structural features of each reconstruction. SAXS bead models were converted to volumetric Situs format with the pdb2vol kernel convolution utility (26).

Results

Rv1364c encodes predicted regulatory, phosphatase, kinase and substrate domains--Four domains showing homology to known folds were identified within the Rv1364c sequence (Fig. 2A;

(18)). The N-terminal domain, from amino acids 1-397, is an RsbU-like phosphatase domain. The RsbU domain is composed of an N-terminal regulatory PAS domain from residues 1-137 followed by a PP2C Ser/Thr phosphatase domain from residues 138-397. Based on similarities to anti-sigma factors, we hypothesize that amino acids 398-544 encode a domain like RsbT that binds and activates the phosphatase and exhibits Ser or Thr kinase activity. The C-terminal domain (residues 545-653) exhibits homology to anti-anti-sigma factors, including a conserved phospho-acceptor site at Ser600. This homology suggests that the C-terminal domain acts like RsbS to bind in an unphosphorylated state to the anti-sigma factor domain and that phosphorylation may trigger phosphatase activation.

Rv1364c RsbT domain activates the phosphatase domain--To test the prediction that Rv1364c encodes an active phosphatase, the protein was expressed and purified from *E. coli* and the kinetics of phosphate ester hydrolysis were measured. Using the noncognate, colorimetric substrate, p-nitrophenylphosphate (pNPP), phosphatase activity was observed for full length Rv1364c (Fig. 2). Either manganese or magnesium was required for hydrolysis, and activity was maximal with both divalent cations present (data not shown). The pH dependence of this activity was assessed using a base-quenched endpoint assay, and the upper bound of accessible pH was limited by the solubility of manganese (Fig. 2*B*). While activity was maximal at pH 8.0, there was only a modest (~6 fold) reduction in activity at pH 5.5, as observed for other PP2C Ser/Thr phosphatases (27).

To assess the potential for interdomain communication, we removed multiple domains from full-length Rv1364c and kinetic constants were measured. At the optimum divalent metal concentrations and pH, the phosphatase activity of full-length Rv1364c was compared to that of a PAS domain truncation (ΔPAS), RsbS domain truncation (ΔRsbS), double PAS and RsbS domain truncation (ΔPASΔRsbS) or PP2C phosphatase domain alone (ΔPASΔRsbSΔRsbT) (Fig. 2*C*). The truncations that retained the PP2C and RsbT domains were fully, or nearly fully active. By comparison, the isolated phosphatase domain alone exhibited reduced k_{cat} and increased Km values, which corresponded to an apparent 9-18-fold activation by the RsbT domain (Table 1). The shortest fully active combination, which was slightly more active than full-length phosphatase, was the PP2C-RsbT fusion (ΔPASΔRsbS). These results show that the anti-sigma-factor (RsbT) domain activates the phosphatase domain through a direct interaction, while the PAS and RsbS domains are dispensable for pNPP hydrolysis *in vitro*. Activation of the phosphatase domain by the anti-sigma factor domain is consistent with the RsbU-RsbT interaction paradigm (Fig. 1; (28)).

Rv1364c RsbT domain encodes an active kinase--RsbT proteins, in addition to binding and activating RsbU phosphatases, have been shown to catalyze transfer of the γ-phosphate from ATP to a serine residue of RsbS proteins (28). To assess the potential kinase activity of the Rv1364c RsbT domain, the phosphorylation states of the constructs were characterized. After expression in *E. coli*, Rv1364c mutants were purified by immobilized-metal-affinity chromatography (IMAC) using Ni^{2+}-Sepharose and analyzed by SDS-PAGE. The variants were expressed at similar levels except for the deletion mutant lacking the RsbT and RsbS domains (ΔRsbTΔRsbS) (Fig. 3*A*, top). ΔRsbTΔRsbS was difficult to express and purify and generally insoluble. Staining these proteins for the presence of phosphoryl groups using Diamond Pro-Q showed that the Rv1364c construct lacking the PAS and phosphatase domains (ΔPAS ΔPhos) was efficiently phosphorylated (Fig. 3*B*, bottom). In contrast, the constructs containing the phosphatase domain or constructs lacking the RsbS domain were phosphorylated at background levels.

To define the basis for the kinase activity, we characterized the effects of mutations in the predicted phospho-acceptor and kinase active sites (Fig. 3*C*). Phosphorylation of the RsbT-RsbS construct (ΔPAS ΔPhos) was abolished by the Ser600Ala mutation of the predicted phosphorylation site in the RsbS domain. Further, the mutations Glu444Lys, Asn448Lys or His452Ala in the predicted kinase active site in the RsbT domain also eliminated phosphorylation.

The Ser600Ala RsbT-RsbS (ΔPAS ΔPhos) mutant, containing a wild-type RsbT kinase domain, did not efficiently phosphorylate the His452Ala mutant containing a wild-type phospho-acceptor site but lacking kinase activity, consistent with the conclusion that the kinase functions in an intramolecular reaction. These results suggest that the RsbT domain functions as a classic anti-sigma-factor-like kinase to phosphorylate the RsbS domain at the canonical regulatory site, and the phosphatase domain antagonizes this activity.

Corroborating this conclusion, the full length Rv1364c exhibited a lower level of phosphorylation than the RsbT-RsbS fragment. Using the phosphatase domain alone (ΔPASΔRsbSΔRsbT) as a negative control, full length (FL) Rv1364c exhibited 11% phosphorylation compared to the RsbT-RsbS fusion (ΔPAS ΔPP2C) (Fig. 3*B*). This apparently incomplete phosphorylation of full-length Rv1364c was observed (Fig. 3*A*) when the proteins were expressed in the ATP-rich environment of *E. coli* for 24 hours and purified in divalent-cation-free buffer in five hours. On the other hand, including magnesium in the buffers during a two-day, multi-column purification in the absence of ATP eliminated phosphorylation of full-length Rv1364c. This protein exhibited ~1% phosphorylation (Fig. 3*B*), which, within experimental error, indicated the absence of phosphate modifications. In contrast, the Asp328Ala mutation in full-length Rv1364c, which inactivates the phosphatase domain by eliminating a conserved carboxylate predicted in PP2C enzymes to coordinate the hydroxyl nucleophile, led to the accumulation of the phosphorylated protein (Fig. 3*C*). These results reflect a dynamic antagonism between the PP2C and kinase domains in Rv1364c.

Small angle X-ray scattering of full-length Rv1364c—The ability to isolate the homogeneously dephosphorylated wild-type protein and the phosphorylated Asp328Ala mutant afforded the opportunity to use SAXS to identify key structural features of the Rv1364c protein responsive to phosphorylation. The protein samples used for SAXS were purified in the presence of magnesium over three columns (IMAC, gel filtration and anion exchange chromatography) and analyzed using dynamic light scattering to ensure the lack of aggregation.

The SAXS profiles of both phosphorylated and unphosphorylated Rv1364c in a concentration range of 1-4 mg/ml exhibited features of a well-behaved protein (data not shown), but scattering in the low-angle region indicated particle interference at high concentrations. Consequently, we determined the interference-free SAXS profile by extrapolating the measured scattering curves to infinite dilution (Fig. 4*A*). Linear Guinier plots (29) of the interference-free profiles (Fig. 4*A*, inset) indicated that the protein adopts a distinct aggregation-free state. The molecular weight of the unphosphorylated protein calculated from these SAXS data using the Porod volume (30) was 150 kDa. The calculated molecular weight of the Rv1364c construct used in this study was 69.7 kDa, so the mass of a dimer (139.4 kDa) was within the 20% standard error associated with such calculations (20). The existence of a dimer in solution also was observed by gel filtration chromatography (Table 2). The radius of gyration (R_G) of the unphosphorylated form calculated from the SAXS data was ~51 Å, which is unusually large for a protein of this size. These results suggest that Rv1364c contains extended elements or adopts a hollow conformation.

Comparison of the X-ray scattering profiles from phosphorylated (Asp328Ala) and unphosphorylated Rv1364c revealed differences over the entire scattering range in the Kratky plot (Fig. 4*B*). The phospho-protein showed broadening of the distribution function, P(r) (Fig. 4*C*), and an increase in R_G from ~51 to ~54 Å. These results provide evidence for a change in the overall shape of the dimer in response to phosphorylation of Ser600.

We determined solution structural envelopes of full-length Rv1364c to generate models for higher-order complex formation between RsbU-, RsbT-, and RsbS-like domains. The P(r) functions calculated by GNOM (23) (Fig. 4*B*), marked by a broad P(r) maximum, indicated that Rv1364c adopts an elongated conformation with maximal dimension (D_{max}) of ~190 Å. Using the program GASBOR (24) to fit the interference-free data in the q range of 0.01–0.30 $Å^{-1}$, we calculated

envelopes constrained by P2 symmetry for both the unphosphorylated (wild-type) and phosphorylated (Asp328Ala) proteins. The commonalities among individual reconstructions of the SAXS envelopes reflect the similarities between independent runs (Fig. 4*D*). To insure the symmetry operator was not inappropriately biasing our calculations, the overall shape was re-calculated without a symmetry operator (Fig. 4*D*), and similar structural features were observed.

Discussion

This report describes and quantifies the overall structure and regulation of Rv1364c, the RsbU-like, environmental Ser/Thr phosphatase of *Mtb*. The functional equivalence of Rv1364c and RsbU and its regulators in *B. subtilis* is supported by previous reports implicating Rv1364c in the *Mtb* sigma-factor-F pathway (18). As bacterial genomes often encode functionally related proteins in close proximity, it is also relevant to note that *Rv1364c* neighbors *rsfA*, the gene for an anti-anti-sigma factor that regulates sigma factor F in response to cellular redox levels (31). Homologies alone, however, are not sufficient to distinguish the disparate functions of such domains. Some anti-sigma factor proteins, for example, bind and inactivate sigma factors while others bind and activate a Ser/Thr phosphatase. The physical connection between the RsbU-like phosphatase and anti-sigma factor/anti-anti-sigma factor pair in Rv1364c suggested a likely interaction, providing clues to domain function and making biochemical studies tractable.

The full-length Rv1364c protein showed pNPP hydrolysis activity *in vitro*, consistent with previous results (18). In addition, we found that fusion to the RsbT domain activated the phosphatase domain ~18-fold (Fig. 2*C*), consistent with a direct regulatory interaction. Neither the PAS domain nor the RsbS domain was essential for this stimulation. The apparently dispensable roles for the terminal domains *in vitro* should not be extrapolated to assume similarly dispensable roles *in vivo*. PAS domains, for example, are thought to bind small-molecule cofactors and interact with other proteins. While dephosphorylated *B. subtilis* RsbS has been shown to bind RsbT, effectively blocking the RsbU-stimulating activity of the RsbT, this RsbS-RsbT binding is significantly diminished in the absence of RsbR and stressosome formation (32). Consistent with this essential role for RsbR, the RsbS domain of Rv1364c did not antagonize RsbT-domain-dependent phosphatase stimulation in the *in vitro* assays (Fig. 2*C*). In the absence of a PAS domain ligand or the intact stressosome, the roles of these terminal domains cannot be fully assessed.

In addition to activating the phosphatase domain, the Rv1364c RsbT domain phosphorylated the RsbS domain. This robust activity (Fig. 3*A*) contrasts with the recent report that the RsbT domain of Rv1364c lacks kinase activity (18). This apparent discrepancy likely results from the use in the previous study of the full-length (active) phosphatase, isolated single domains or noncognate substrates to assay for phospho-transfer. In addition, the previous *in vitro* kinase assays employed low concentrations (8 μM) of Mg^{2+}-ATP (18), rather than physiological concentrations in the mM range. We found that the kinase activity is antagonized in the full-length protein by the phosphatase domain (Fig. 3), which limits the accumulation of phosphorylated product. The Asp328Ala mutation in the phosphatase active site led to the accumulation of phosphorylated full-length protein (Fig. 3*C*), consistent with the idea that the interplay of kinase- and phosphatase-domain activities controls the phosphorylation state. Deletion or inactivation of the phosphatase domain unmasked the kinase activity.

The previously observed lack of kinase activity also was based in part on phosphorylation assays using the noncognate substrates, myelin basic protein and histones, which lack competent recognition sites (18). In addition, phosphorylation of the Rv1364c RsbS-like anti-anti-sigma factor domain was tested only in trans using the isolated domains (18). Consistent with these intermolecular kinase assays, we found that the Ser600Ala RsbT-RsbS construct did not phosphorylate the His352Ala mutant lacking a functional kinase (Fig. 3*C*). In contrast, the Rv1364c RsbT-RsbS domain fusion showed stable phosphorylation in *E. coli*, consistent with the idea that this segment contains both the kinase domain and the site of intramolecular phosphorylation.

Homologies to active anti-sigma-factor kinases suggested that conserved residues Glu444, Asn448 and His452 in Rv1364c are located in the ATP binding site. Similarly, Ser600 in the Rv1364c RsbS domain matches the phospho-acceptor residues in sequence alignments with functionally phosphorylated anti-anti-sigma factor proteins. Mutations of either the predicted kinase active site or the phospho-acceptor site blocked phosphorylation of the RsbT-RsbS construct (ΔPASΔPP2C; Fig. 3*C*). Rather than requiring an entirely new mechanism of regulation in *Mtb* (18), these results suggest that the RsbT-like kinase domain of *Mtb* Rv1364c fits the paradigm established in *B. subtilis* (13) for regulation of the environmental phosphatase.

Diminished phosphorylation was observed in the constructs containing the Rv1364c phosphatase domain, and the intact protein purified from *E. coli* was partially phosphorylated (Fig. 3). Incubation of partially phosphorylated full-length protein with Mg^{2+} and Mn^{2+} removed the remaining phosphates. We therefore conclude that the phosphatase domain of Rv1364c acts on the same substrate as the kinase domain. It remains to be determined if this reaction is stimulated by other components of the stressosome, as observed in *B. subtilis*.

To characterize the structural consequences of phosphorylation, we used SAXS to determine the overall shape of Rv1364c and the changes in domain organization in response to phosphorylation. The SAXS data in this report characterize, for the first time, the interactions between RsbU-, RsbT-, and RsbS-like domains. Unlike other bacteria in which unphosphorylated RsbS is expected to sequester RsbT and block binding to the RsbU phosphatase, these domains are held together in *Mtb* by fusion into the single Rv1364c polypeptide chain. Despite differences in the X-ray scattering curves, the SAXS patterns demonstrate that both the phosphorylated and unphosphorylated *Mtb* Rv1364c proteins form extended dimers. The radii of gyration, 51 Å for the unphosphorylated protein and 54 Å for the Asp328Ala phosphoprotein, are large for a dimer of this molecular mass, reflecting the presence of arms extended from the central globular core (Fig. 4*D*). Both dimers are relatively flat, with a possible reduction in density around the 2-fold rotation axis.

Attempts to fit high-resolution crystallographic models of homologs of the individual domains into the SAXS envelope were unsuccessful. This difficulty likely has multiple causes, including the sequence divergence of the homologs with structures in the Protein Data Bank (PDB). The PAS domain, for example, exhibited only 17% homology to the closest homolog in the PDB. Additional challenges to fitting atomic models into the SAXS envelopes also may be presented by the possible flexibility and conformational disorder of the extended regions, the low resolution of the envelopes, uncertainty about the location of the dimer interface and potential heterogeneity due to differences in aggregation, phosphorylation and conformational switching.

Valuable information about the structure, however, can be gleaned from the combination of gel filtration elution profiles, SAXS envelopes, and biochemical assays. All of the most active phosphatase constructs were multimers in solution (Table 2 and Fig. 2*C*). Only the phosphatase domain in isolation was a monomer, and this monomeric RsbU-like domain showed 18-fold lower catalytic activity compared to the dimeric RsbU-RsbT domain fusion. While the presence of a short coiled-coil sequence has been predicted in the phosphatase segment (18), this region contains many noncanonical predicted interface residues (33) and clearly is insufficient to mediate association (Table 2). Fusion of the RsbT anti-sigma-factor domain to the RsbU domain was sufficient to mediate dimerization. This association parallels the finding that a number of anti-sigma factors from different species form dimers in isolation (16,34,35). The SAXS envelope of full-length Rv1364c not only confirmed the dimeric arrangement, but also placed limits on the location of the three domains. An arrangement in which the anti-sigma factor domain dimerizes and interacts with the phosphatase domain of the opposing monomer in the central globular region of this envelope would maintain the physical interactions necessary to account for the biochemical activity and the

structural data. Notably, the phosphorylation of Rv1364c results in opening of the central region of the dimer (Fig. 4). This conformational change may alter the arrangement of the RsbU-, RsbT-, and RsbS-like domains and provides evidence for phosphorylation-dependent partner switching of domains that ultimately controls the bacterial transcriptional response.

Higher-order complex formation and partner switching play essential roles in sigma-factor regulatory cascades. The ability of the Rv1364c anti-sigma-factor kinase domain to stimulate phosphatase activity and to phosphorylate the RsbS domain in an intramolecular reaction suggests that this multi-domain protein provides a tractable system to characterize the general features of environmental phosphatase activity and regulation across the bacterial kingdom.

Acknowledgements

We thank Ho-Leung Ng for advice regarding SAXS data collection as well as Jacob Corn and the scientists at the SIBYLS beamline for support with SAXS data collection and processing. We thank E. Megan Flynn and Christoph Grundner for advice regarding kinetic analysis, and James Fraser for network assistance. The Office of Biological and Environmental Research, U.S. Department of Energy, under contract DE-AC02-05CH11231 supports the SIBLYS beamline (ALS BL12.3.1), and Laboratory Research Computing at LBNL supports the computational analysis of SAXS data. This work was sponsored by a National Science Foundation Graduate Research Fellowship (A.G.) and the TB Structural Genomics Consortium (NIH grant P01 AI68135 to T.A.).

32. Chen, C. C., Lewis, R. J., Harris, R., Yudkin, M. D., and Delumeau, O. (2003) *Mol Microbiol* **49**, 1657-1669
33. Woolfson, D. N., and Alber, T. (1995) *Protein Sci* **4**, 1596-1607
34. Delumeau, O., Lewis, R. J., and Yudkin, M. D. (2002) *J Bacteriol* **184**, 5583-5589
35. Urbauer, J. L., Simeonov, M. F., Urbauer, R. J., Adelman, K., Gilmore, J. M., and Brody, E. N. (2002) *Proc Natl Acad Sci U S A* **99**, 1831-1835
36. Kim, T. J., Gaidenko, T. A., and Price, C. W. (2004) *J Mol Biol* **341**, 135-150
37. Vijay, K., Brody, M. S., Fredlund, E., and Price, C. W. (2000) *Mol Microbiol* **35**, 180-188

Table 1. Apparent kinetic constants for pNPP hydrolysis of Rv1364c variants. Km, Vmax, and kcat were calculated by nonlinear regression of the data (e.g. Fig. 2*C*) using SigmaPlot. The unphosphorylated fusions containing the PP2C and RsbT (anti-sigma factor) domains showed the highest activation relative to the isolated PP2C domain (last column).

Domains present	Vmax (μM/min)	Km (mM)	kcat (/min)	kcat/Km (mM/min)	Fold activation
All (full-length)	12.2 +/- 0.8	19.1 +/- 2.9	14.2 +/- 0.93	0.74	9.3
PP2C-RsbT	15.7 +/- 0.7	13.3 +/- 1.5	18.3 +/- 0.8	1.4	17.5
PP2C	2.1 +/- 0.5	30.6 +/- 2.5	2.4 +/- 0.6	0.08	1
PP2C-RsbT-RsbS	15.6 +/- 0.5	14.5 +/- 1.1	18.1 +/- 0.6	1.2	15
PAS-PP2C-RsbT	16.6 +/- 0.7	20.2 +/- 1.7	19.3 +/- 0.8	0.95	11.9

Table 2. Apparent molecular weight, as observed by size exclusion chromatography, of Rv1364c variants. While the phosphatase domain alone was monomeric, all other constructs formed apparent dimers or tetramers in solution.

Domains present	Theoretical MW (kDa)	Observed MW (kDa)	Predicted state
All (full-length)	69.5	148.3	dimer
PP2C-RsbT	42.9	91.4	dimer
PP2C	27.0	30.6	monomer
PP2C-RsbT-RsbS	54.0	213.8	tetramer
PAS-PP2C-RsbT	58.4	245.7	tetramer

Figure 1

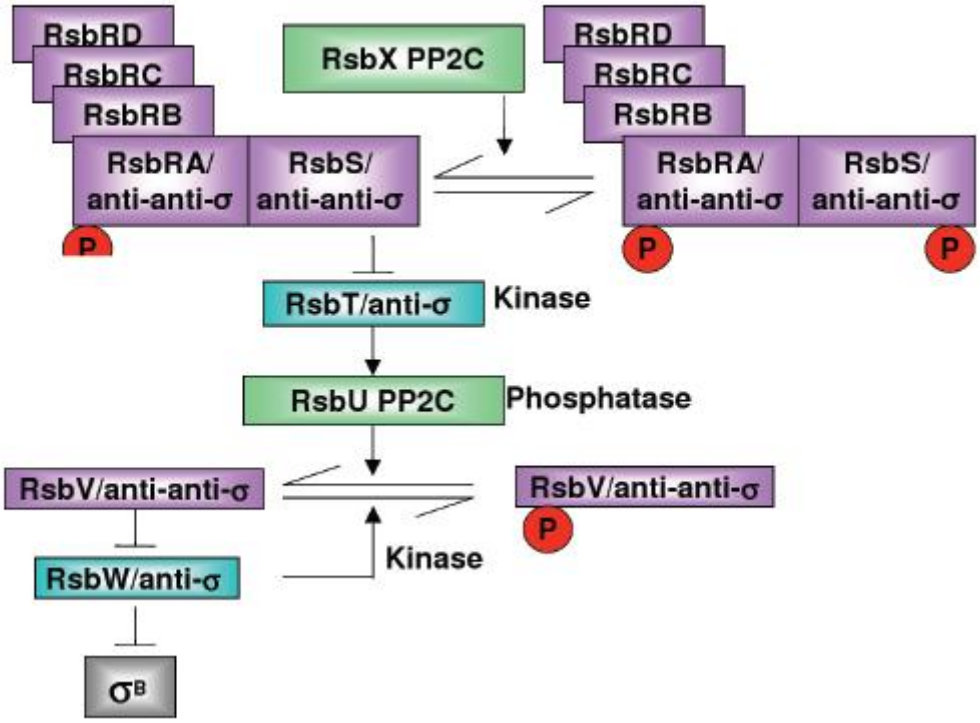


Figure 1. Model of alternative sigma factor regulation in *B. subtilis*.
RsbW, an anti-sigma factor (blue), binds and inactivates sigma factor B (gray). RsbV, an anti-anti sigma factor (purple), binds and inactivates RsbW, functionally relieving sigma factor B inhibition. RsbW phosphorylates RsbV, which dissociates and allows RsbW to bind sigma factor B. The RsbU phosphatase, with an N-terminal PAS domain and C-terminal PP2C phosphatase domain, is the *B. subtilis* homolog of *Mtb* Rv1634c. RsbU dephosphorylates RsbV, allowing it to again bind RsbW. RsbU is activated by binding to RsbT, a protein with sequence homology to RsbW (blue). RsbS (purple) binds and sequesters RsbT in the high-molecular-weight RsbR-RsbS "stressosome" complex (purple). Despite the distinct roles of RsbT and RsbW, they are homologous at the sequence level. The RsbR proteins, RsbV and RsbS are also paralogs. Adapted from (36).

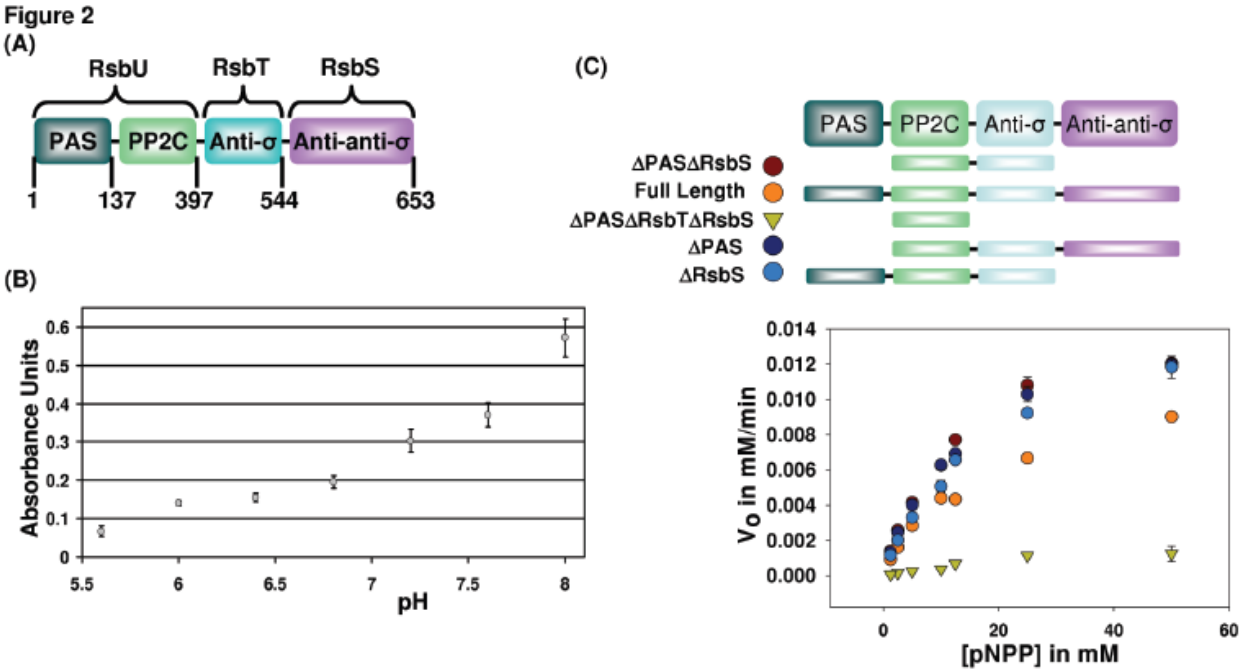


Figure 2. ***M. tuberculosis*** **Rv1364c encodes an active phosphatase domain.**
(A) Rv1364c contains four distinct domains. The N-terminal PAS domain is thought to sense energy levels in the cell and regulate the phosphatase (37). Residues 138-397 encode a PP2C phosphatase domain, and the N-terminal 397 residues comprise a RsbU-like domain. Residues 398-544 comprise an anti-sigma factor domain, which we hypothesize functions like RsbT. The C-terminal domain, which we hypothesize functions like RsbS to bind the RsbT domain, exhibits homology to anti-anti-sigma factors.
(B) Dependence of p-nitrophenylphosphate (pNPP) hydrolysis by full-length Rv1364c protein as a function of pH. The pH dependence was shallow (18), and activity was greatest at the highest pH tested (8.0).
(C) Kinetic analysis of full-length and domain constructs of Rv1364c (top) using pNPP as the substrate. The PP2C-RsbT-domain combination (maroon octagons) comprises the minimal functional unit for full phosphatase activity toward pNPP. Consistent with the hypothesis that the anti-sigma factor domain functions as an RsbT-like phosphatase activator, the PP2C domain alone (green triangles) is significantly less active than the PP2C-RsbT domain combination.

Figure 3

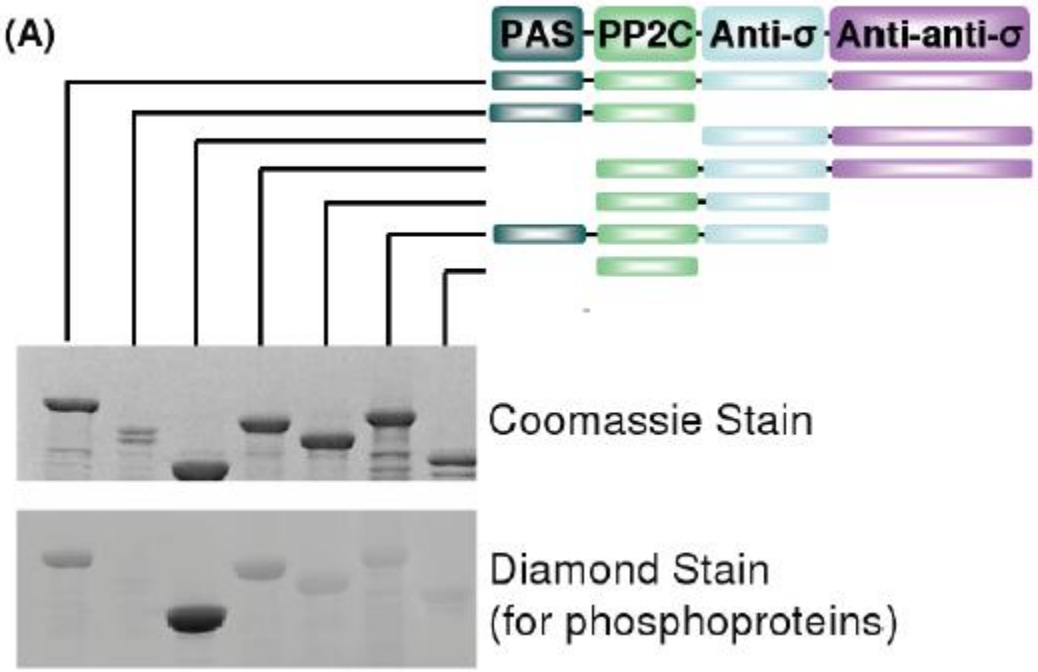
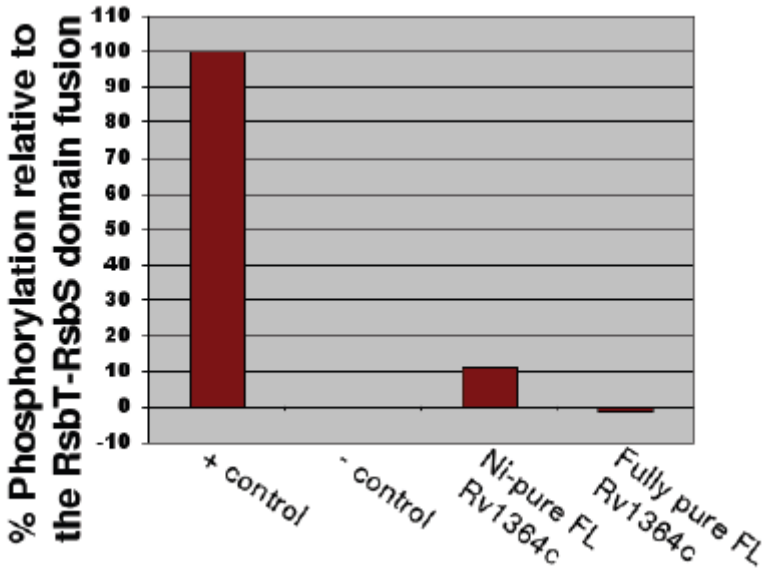


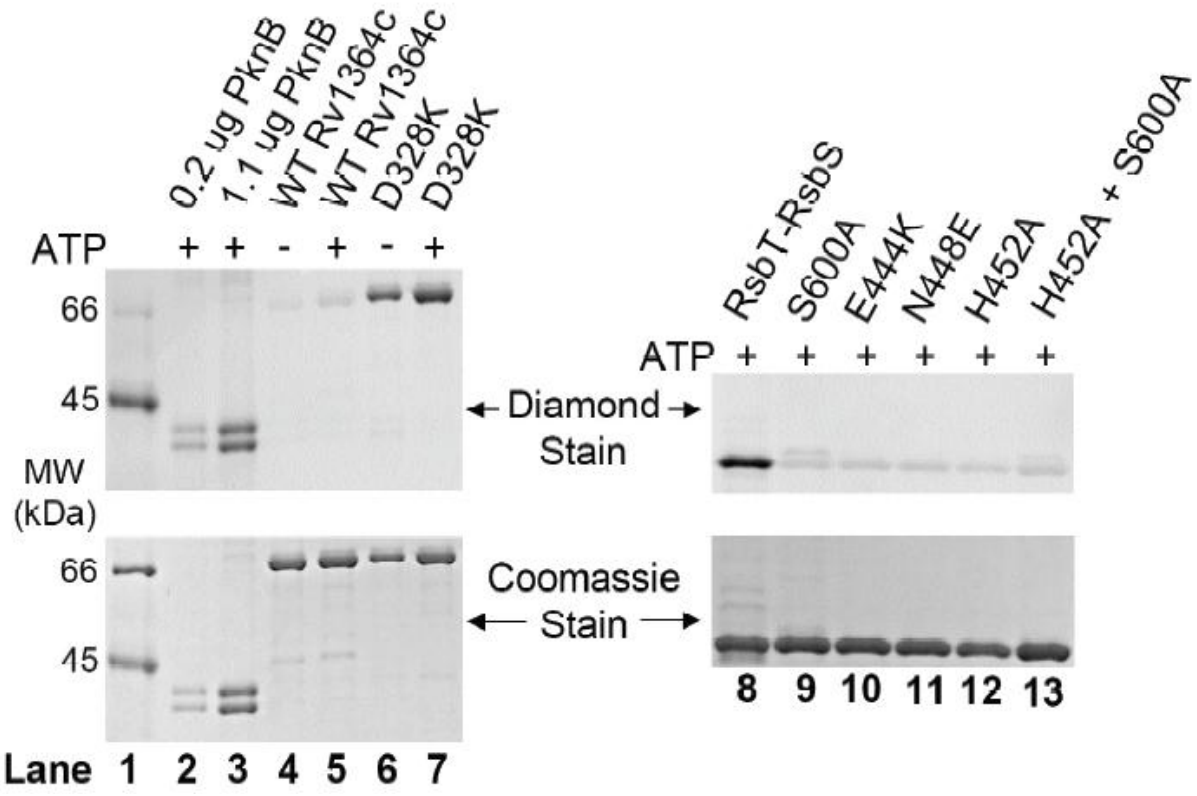
Figure 3. The phosphatase domain antagonizes Rv1364c autophosphorylation.
(A) Rv1364c contains active phosphatase and kinase domains. Purified domain-truncation mutants of Rv1364c were stained with Coomassie blue to measure protein levels (top) or Diamond Pro-Q, which preferentially stains phosphoproteins (bottom). The ΔRsbTΔRsbS mutant was unstable during expression and purification, and this construct is represented by less total protein. The ΔPASΔPP2C variant (lane 3), containing the predicted anti-sigma factor kinase domain and the phospho-acceptor in the anti-anti-sigma factor domain, is significantly phosphorylated after expression in *E. coli*. While the PP2C domain alone exhibited a typical level of background Diamond staining associated with non-phosphorylated proteins, the full-length protein appeared slightly phosphorylated when purified using a single-column, one-day protocol.

Figure 3(B)



(B) Phosphorylation levels observed by Diamond staining were quantified for ΔPASΔPP2C (positive control), PP2C domain alone (negative control), full-length Rv1364c purified rapidly as in (A), and Rv1364c after more extensive purification in the presence of magnesium for two days. Rv1364c, after the one-day purification, stained at 11% of the positive control, while the same protein treated with magnesium stained at -1% of the positive control.

Figure 3(C)



(C) Effects of amino-acid substitutions in the kinase, phosphatase, or phospho-acceptor sites on the phosphorylation state of Rv1364c. Asp328 is predicted to coordinate the catalytic hydroxyl nucleophile in the RsbU-like phosphatase domain; Glu444, Asn 448 and His452 are predicted to contact ATP in the RsbT-like kinase domain; and Ser600 matches the phospho-acceptor in RsbS-like anti-anti-sigma factors. SDS gel stained to detect phosphorylation (Pro-Q Diamond, top) and protein levels (Coomassie Blue, bottom). Lanes 1-3 contain the Peppermint Stick molecular mass markers and *Mtb* PknB positive controls for phosphoprotein staining. Lanes 4-7 contain full-length Rv1364c, and lanes 8-13 contain the

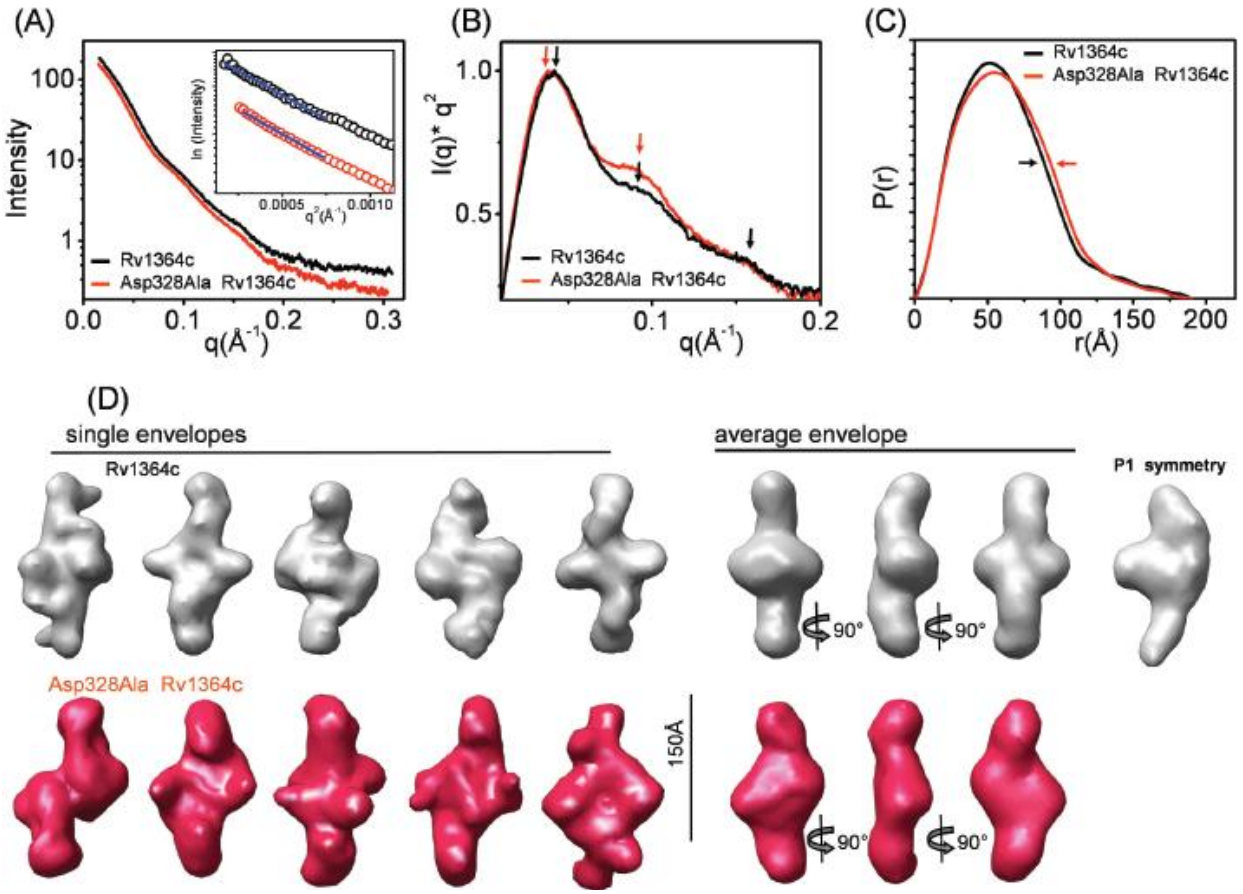


Figure 4. Small angle x-ray scattering (SAXS) reveals key Rv1364c structural features and a phosphorylation-dependent conformational change.

(A) The interference-free experimental scattering curve for full-length Rv1364c (black) and the phosphorylated Asp328Ala mutant (red). A Guinier plot (inset) with the linear fit (blue line) within the limits qR_G>1.3 was used to calculate R_G values of 51 Å for the unphosphorylated protein and 54 Å for the phosphorylated mutant.

(B) SAXS profiles shown in the Kratky plot ($I(q)*q^2$ vs. q^2). Differences in the Kratky plots highlighted by black and red arrows indicate phosphorylation-induced conformational changes.

(C) The pair distribution function, P(r), indicates an elongated particle, with a maximum dimension, D_{max}, of ~190 Å, and broadening of the structure (highlighted with arrows) consistent with reorientation of the domains upon phosphorylation.

(D) Five representative models of unphosphorylated (gray, top) and phosphorylated (red, bottom) Rv1364c reconstructed in P2 symmetry using GASBOR (24) are shown in a surface representation calculated using the SITUS package (26). Consistency between the independently generated models was high, with χ^2 values of 1.2-1.5. The averaged envelopes show flat extended dimers with a globular core. Global differences between the models of the unphosphorylated and phosphorylated proteins provide evidence for partner switching of the domains. A representative view of the GASBOR average model reconstructed with a P1 symmetry operator (right) shows a similar overall shape to models constrained by the P2 operator.

References

1. Zhang, Y. (2004) *Front Biosci* **9**, 1136-1156
2. Geiman, D. E., Kaushal, D., Ko, C., Tyagi, S., Manabe, Y. C., Schroeder, B. G., Fleischmann, R. D., Morrison, N. E., Converse, P. J., Chen, P., and Bishai, W. R. (2004) *Infect Immun* **72**, 1733-1745
3. Song, T., Dove, S. L., Lee, K. H., and Husson, R. N. (2003) *Mol Microbiol* **50**, 949-959
4. Hampshire, T., Soneji, S., Bacon, J., James, B. W., Hinds, J., Laing, K., Stabler, R. A., Marsh, P. D., and Butcher, P. D. (2004) *Tuberculosis (Edinb)* **84**, 228-238
5. Greenstein, A. E., MacGurn, J. A., Baer, C. E., Falick, A. M., Cox, J. S., and Alber, T. (2007) *PLoS Pathogens* **3**, e49
6. Said-Salim, B., Mostowy, S., Kristof, A. S., and Behr, M. A. (2006) *Mol Microbiol* **62**, 1251-1263
7. Rodrigue, S., Provvedi, R., Jacques, P. E., Gaudreau, L., and Manganelli, R. (2006) *FEMS Microbiol Rev* **30**, 926-941
8. Campbell, E. A., Westblade, L. F., and Darst, S. A. (2008) *Curr Opin Microbiol* **11**, 121-127
9. Hughes, K. T., and Mathee, K. (1998) *Annu Rev Microbiol* **52**, 231-286
10. Bashyam, M. D., and Hasnain, S. E. (2004) *Infect Genet Evol* **4**, 301-308
11. Clarkson, J., Campbell, I. D., and Yudkin, M. D. (2004) *J Mol Biol* **342**, 1187-1195
12. Britton, R. A., Eichenberger, P., Gonzalez-Pastor, J. E., Fawcett, P., Monson, R., Losick, R., and Grossman, A. D. (2002) *J Bacteriol* **184**, 4881-4890
13. Kim, T. J., Gaidenko, T. A., and Price, C. W. (2004) *J Bacteriol* **186**, 6124-6132
14. Pane-Farre, J., Lewis, R. J., and Stulke, J. (2005) *J Mol Microbiol Biotechnol* **9**, 65-76
15. Chen, C. C., Yudkin, M. D., and Delumeau, O. (2004) *J Bacteriol* **186**, 6830-6836
16. Parida, B. K., Douglas, T., Nino, C., and Dhandayuthapani, S. (2005) *Tuberculosis (Edinb)* **85**, 347-355
17. Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999) *Science* **285**, 751-753
18. Sachdeva, P., Narayan, A., Misra, R., Brahmachari, V., and Singh, Y. (2008) *Febs J* **275**, 6295-6308
19. Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., 2nd, Tsutakawa, S. E., Jenney, F. E., Jr., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S. J., Scott, J. W., Dillard, B. D., Adams, M. W., and Tainer, J. A. (2009) *Nature methods* **6**(8), 606-612
20. Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007) *Quarterly reviews of biophysics* **40**(3), 191-285
21. Bennett-Lovsey, R. M., Herbert, A. D., Sternberg, M. J., and Kelley, L. A. (2008) *Proteins* **70**, 611-625
22. Busso, D., Delagoutte-Busso, B., and Moras, D. (2005) *Anal Biochem* **343**, 313-321
23. Svergun, D. (1992) *J. Appl. Cryst.* **25**, 495-503
24. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001) *Biophys. J.* **76**, 2879-2886
25. Volkov, V. V., and Svergun, D. I. (2003) *J. Appl. Cryst.* **36**, 860-864
26. Wriggers, W., Milligan, R. A., and McCammon, J. A. (1999) *Journal of structural biology* **125**, 185-195
27. Fjeld, C. C., and Denu, J. M. (1999) *J Biol Chem* **274**, 20336-20343
28. Yang, X., Kang, C. M., Brody, M. S., and Price, C. W. (1996) *Genes Dev* **10**, 2265-2275
29. Guinier, A., and Fournet, F. (1955) *Small angle scattering of X-rays*, Wiley Interscience, New York
30. Porod, G. (1982) General Theory. in *Small Angle X-ray Scattering* (Glatter, O., and Kratky, O. eds.), Academic Press, London. pp 17-51
31. Beaucher, J., Rodrigue, S., Jacques, P. E., Smith, I., Brzezinski, R., and Gaudreau, L. (2002) *Mol Microbiol* **45**, 1527-1540

32. Chen, C. C., Lewis, R. J., Harris, R., Yudkin, M. D., and Delumeau, O. (2003) *Mol Microbiol* **49**, 1657-1669
33. Woolfson, D. N., and Alber, T. (1995) *Protein Sci* **4**, 1596-1607
34. Delumeau, O., Lewis, R. J., and Yudkin, M. D. (2002) *J Bacteriol* **184**, 5583-5589
35. Urbauer, J. L., Simeonov, M. F., Urbauer, R. J., Adelman, K., Gilmore, J. M., and Brody, E. N. (2002) *Proc Natl Acad Sci U S A* **99**, 1831-1835
36. Kim, T. J., Gaidenko, T. A., and Price, C. W. (2004) *J Mol Biol* **341**, 135-150
37. Vijay, K., Brody, M. S., Fredlund, E., and Price, C. W. (2000) *Mol Microbiol* **35**, 180-188