Experimental Characterization of *Mycobacterium tuberculosis* Adenosine Nucleotide Binding and Ser/Thr/Tyr Phosphosignaling

Corrie Maria Ortega

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Christoph Grundner, Chair

David R. Sherman

Colin Manoil

Program Authorized to Offer Degree:

Pathobiology

University of Washington

Abstract

Experimental Characterization of *Mycobacterium tuberculosis*

Adenosine Nucleotide Binding and Ser/Thr/Tyr Phosphosignaling

Corrie Maria Ortega

Chair of the Supervisory Committee:

Christoph Grundner, Ph.D.

Pathobiology Program, Department of Global Health

Tuberculosis (TB) is responsible for 1.3 million deaths each year and is a global health burden of overwhelming proportions. The success of *Mycobacterium tuberculosis (Mtb)*, the causative agent of TB, lies in part with the pathogen's ability to persist under growth-limiting stress to cause disease months to years later. With drug resistant strains on the rise, the need for novel therapeutics is urgent. A fundamental problem for understanding *Mtb* metabolism and pathogenesis and for the development of novel therapeutics is the lack of reliable annotation for many *Mtb* proteins. In addition, our lack of understanding the molecular basis of persistence further hampers the development of effective therapies. Functional characterization of the proteome and the molecular networks that drive *Mtb* persistence is critical to inform novel therapeutic approaches. This dissertation will detail efforts to: 1. Annotate the *Mtb* ATP binding proteins (ATPome), 2. Characterize the role of serine/threonine protein kinases in growth-limiting adaptions that underlie persistence, and 3. Identify protein Tyr phosphorylation, a new *Mtb* phosphosignaling mechanism.

The majority of functional annotation is based on in silico predictions; however, these inferences are often unreliable. In addition, ~25% of the *Mtb* protein-coding genes have no predicted function and are annotated as hypothetical proteins. To address the lack of functional annotation, we used a high-throughput quantitative activity-based protein profiling (ABPP) platform to probe, annotate, and validate the large family of ATP-binding proteins. We experimentally confirmed ~240 in silico predictions and identified 72 hypotheticals as ATP-binding proteins, a number of which are only found in mycobacteria and essential for *in vitro* growth or infection. These data help further define protein function in the *Mtb* proteome and show the diversity among ATP-binding proteins, which may be harnessed for the development of therapeutics.

Mtb pathogenesis is characterized by adaptation to growth-limiting stress. The molecular mechanisms underlying adaptation are largely unknown; however, phosphosignaling has been suggested as a key mediator of stress response. To test this idea, we investigated a subset of ATP-binding proteins, the 11 eukaryotic-like serine/threonine protein kinases, to determine their role in mediating *Mtb* persistence under hypoxia, a growth-limiting stress induced by the host. *Mtb* adapts to hypoxia by entering reversible bacteriostasis until ample oxygen levels return and growth resumes. We found an essential serine/threonine protein kinase, PknB, to be a critical mediator of oxygen-dependent replication. Inhibition of PknB activity in aerated culture, and overexpression of PknB in hypoxia compromised *Mtb* viability. These data point to a model in which *Mtb* uses PknB to drive cell growth and division, but depletes PknB under hypoxic conditions to promote bacteriostasis and persistence. With serine/threonine protein kinases established as drug targets, these findings suggest PknB as a target in all stages of the *Mtb* life cycle, including in non-replicating bacilli thought to underlie persistence and latent infection.

Bacteria drive cellular responses and physiological processes with a variety of ATP-dependent signaling molecules, including serine/threonine, histidine/aspartate, and tyrosine kinases. *Mtb* lacks canonical tyrosine kinases and no tyrosine phosphorylation was detected in

previous phosphoproteomic studies, leading to the current notion that *Mtb* does not support protein Tyr phosphorylation. However, protein Tyr phosphorylation is widespread and found in most bacterial phyla, leading us to reevaluate the presence of Tyr phosphorylation in *Mtb*. We show that several serine/threonine protein kinases, including PknB, are in fact dual specificity kinases that are themselves regulated in part by tyrosine phosphorylation, and that *Mtb* supports extensive tyrosine phosphorylation *in vivo*. These data identify a previously unrecognized arm of *Mtb* phosphosignaling and may present new ways to target *Mtb*.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Our experiences are greatly shaped by those around us. To my mentor and committee, thank you for all of your support and guidance during my graduate career. And to my friends and family, thank you for your endless encouragement and kind words.

Chapter 1 – Introduction

Mycobacterium tuberculosis

The majority of species comprising the genus Mycobacteria are non-pathogenic; however, a subset is highly pathogenic, including *Mycobacterium tuberculosis* (*Mtb),* the causative agent of human tuberculosis (TB). Evidence of *Mtb*'s existence dates back ~20,000 years, but it was not until 1882 that Robert Koch first identified *Mtb* as the causative agent of TB [1]. Characteristic features of *Mtb* include its intracellular pathogenesis, slow growth, its ability to enter reversible bacteriostasis, and relative genetic homogeneity. The *Mtb* genome was sequenced in 1998 and consists of ~4.4 million base pairs that contain ~4,000 protein-coding genes [2]. Subsequent comparative genome analyses with non-pathogenic mycobacterial strains and genome-wide mutagenesis screens have identified genes that confer virulence and that are essential for *in vitro* growth and infection [3-5]; however, the functional roles of many of these genes have not been defined, limiting our understanding of *Mtb* pathogenesis. Better functional annotation of the *Mtb* genome and a better understanding of mechanisms underlying *Mtb* pathogenesis, such as its ability to enter reversible bacteriostasis, will aid in targeting this pathogen that has plagued human health for centuries.

Tuberculosis Infection

Mtb is an obligate aerobic bacterium that grows best under high levels of oxygen. It is not surprising then that *Mtb* infection primarily affects the lungs and transmission, as a result, occurs through the respiratory route. Infected individuals release aerosol-containing bacilli that are subsequently breathed in by naive hosts. Once *Mtb* reaches the lung, it is taken up by resident alveolar macrophages, a subset of innate immune cells that control microbial infection through lysosome-mediated killing [6]. *Mtb* resists lysosomal killing by preventing maturation and acidification of the lysosome through virulence factors that modulate host proteins [7]. *Mtb*

replicates until the host cell ruptures and then infects other macrophages recruited to the site of infection. Additional types of immune cells recruited to the site of infection, including neutrophils, T cells, and B cells, control dissemination by surrounding *Mtb*-infected macrophages and forming ordered, granulomatous structures [8], a hallmark of TB infection. Granulomas can vary in size and shape, and occur simultaneously at different sites in the lung [6,9]. The metabolic and replicative state of *Mtb* within the granuloma is unclear. Prevailing opinion is that *Mtb* is phenotypically heterogeneous, with a subset of persistent bacilli that initially escape immune clearance to expand and disseminate months to years later, resulting in acute disease [10]. The integrity of the granulomas and how well they contain *Mtb* is thought to be a key factor in influencing *Mtb*'s phenotypic heterogeneity and clinical disease outcome.

Infection results in a spectrum of asymptomatic and symptomatic disease with varying severity in humans. Approximately 60% of *Mtb* infections result in an asymptomatic, clinically latent disease [11]. Latently infected individuals react positively to immunological tests but do not present overt disease symptoms and do not transmit *Mtb*. Reactivation occurs when latent infection progresses to active disease, leading to increased bacterial growth and dissemination. Active pulmonary disease is characterized by fatigue, fever, and coughing, which facilitates transmission of the bacilli. With an infectious dose between one and six bacilli [6], individuals with active disease are highly contagious and further spread the disease by releasing infectious aerosols that are subsequently inhaled by naïve hosts.

Tuberculosis Epidemiology

There are 22 high TB burden countries that account for 80% of global TB cases [12]. In 2012, the majority of worldwide cases were reported in South-East Asia (29%), Africa (27%) and the Western Pacific (19%) [12]. India and China alone see the highest burden of cases, accounting for 26% and 12% of total cases, respectively [12].

Approximately one third of the global population harbors a latent TB infection, and it is estimated that 2-10% of cases will reactivate and progress to active, symptomatic disease in their lifetime [13-15]. Co-infection with HIV significantly increases TB reactivation, with 2-10% of HIV-infected individuals reactivating each year [12,16,17]. Today, TB is responsible for more 1.3 million deaths each year, and is the leading cause of mortality for HIV infected individuals [12].

Current Therapeutics

Current treatment regimens for latent and active TB often require multiple antibiotics for lengthy periods of time. For latent infection, recommended treatment is either 6 to 9 months with isoniazid (INH), 3 months with INH and rifapentine, or 4 months with rifampin (RIF). For active disease, the recommended treatment is 2 months with INH, RIF, ethambutol, and pyrazinamide, followed by 4 to 5 months of INH and RIF [18]. The bacilli's intrinsic mechanisms for drug tolerance lead to the long treatment times, resulting in non-compliance. Non-compliance, in turn, underlies relapse and the emergence of drug resistant *Mtb* strains. Disease strains resistant to a single antibiotic have been documented in every country surveyed, and the incidence of multiple drug resistant (MDR), extensive drug resistant (XDR), and now totally drug resistant (TDR) *Mtb* is on the rise [12]. Approximately 450,000 people develop MDR-TB each year, and it is estimated that 9.6% of those are actually XDR-TB cases [12]. As drug resistance increases there is a pressing need to develop novel, more efficient TB therapeutics, especially anti-TB drugs that target drug tolerant, persistent *Mtb*.

As a preventative measure, vaccination with the live attenuated vaccine, Bacillus Calmaette-Guèrin or BCG, is administered in endemic regions to prevent childhood tuberculous meningitis and miliary disease; however, the vaccine's protective effect is highly variable depending on geographic location, and it does not protect against the most prevalent form of disease, adult pulmonary TB [19]. Furthermore, because BCG is a live attenuated vaccine,

vaccination is not recommended for HIV-infected individuals, a population most at risk for TB disease and TB-related death [20].

Hypoxia and Reaeration model of Latency and Reactivation

Latency and reactivation shape TB disease. The environmental factors that drive latency transitions are poorly characterized but of these, oxygen tension may be the best understood and is likely a major determinant. Oxygen tension and *Mtb* growth are closely linked in human and animal infections. *Mtb* preferentially establishes infection in the most oxygenated areas [9] of the lung and reactivation occurs in the upper lobe, one of the most aerated regions [21]. Upon infection, a barrage of immune cells surrounds *Mtb*. Many of the granulomas that ensue are depleted of oxygen at their center compared to surrounding lung tissue [22,23]. Early *in vitro* studies demonstrated that *Mtb* arrests growth under low oxygen tension to persist in a viable non-replicating state for months to years [24]. Therefore, it is thought that a subset of *Mtb* enters a state of reversible non-replication to survive growth-limiting stress, and that this quiescent subset underlies the drug tolerance and persistence characteristic of latent infection. Once the granuloma is compromised and regains contact to oxygen, growth can resume.

Based on the close relationship between *Mtb* growth and oxygen, hypoxia has become the primary model to simulate latency *in vitro*. Originally developed by Larry Wayne, the *in vitro* hypoxia model relied on gradual oxygen depletion by *Mtb* in sealed, standing culture [25,26]. Agitation and a defined culture-to-headspace ratio improved reproducibility [27]. Most recently, a highly reproducible, defined culturing system was developed employing rapid oxygen depletion and constant hypoxic environments using fermentor technology [28,29]. In this model, early log phase culture is kept at full aeration for 24 hours, and then oxygen tension is reduced to <1% of normoxia within 30 minutes. The culture is maintained in hypoxia for 7 days and then normal oxygen tension restored.

The hypoxia model provides insight into infection dynamics and *Mtb* adaptive responses that allow for its persistence under host-induced stress. Several studies have shown large, coordinated changes in gene expression in response to hypoxia. In hypoxia, *Mtb* upregulates alternative sigma factors that can promote broad changes in gene expression [30]. Indeed, the initial hypoxic response is characterized by ~100 genes with altered expression, including genes involved in repression of protein synthesis, cell division, lipid or amino acid synthesis, and aerobic metabolism and upregulation of stress response genes [28]. Beyond the initial response to hypoxia, *Mtb* differentially expresses 230 genes termed the enduring hypoxic response [28]. Further investigation to identify mediators of oxygen-dependent replication could uncover proteins critical to *Mtb*'s survival under growth-limiting stress and could provide latency-specific targets to eliminate non-replicating, drug tolerant *Mtb*.

Functional Gene Annotation in *Mtb*

The genome of the *Mtb* lab strain H37Rv was sequenced in 1998. The genome is ~ 4.4 million base pairs and contains approximately 4,000 protein-coding genes [2]. The sequencing of the *Mtb* genome was one of the most impactful scientific advances in TB research. Sequencing data provided the foundation for the identification of essential and virulence-associated genes, and presented the first insights into gene function using homology and genomic context–based methods. Although these in silico functional analyses provided a wealth of knowledge, they are inferences and often unreliable. In addition, approximately 25% of protein-coding genes in the *Mtb* genome are hypothetical [31], making it exceedingly difficult to assign function with in silico approaches. Experimental characterization of protein function is laborious and exceedingly slow, making functional annotation one of the most difficult challenges in the post-genomic era. The lack of functional annotation of the *Mtb* proteome and understanding of the biological roles proteins play in the context of infection provides a significant obstacle at all levels of TB research, including drug development.

To address the current limitation in experimental annotation, we developed a chemical biology platform, activity-based protein profiling (ABPP) combined with quantitative mass spectrometry-based proteomics. ABPP uses active site-directed covalent probes, activity-based probes (ABP), to label entire enzyme families in complex proteomes. ABPs target enzymes with shared catalytic mechanisms, allowing the measurement of enzyme activities. Since enzyme activity can be largely independent from the enzyme's expression or protein abundance, ABPP provides unique functional information. Arguably, ABPP is currently the only, high-throughput quantitative readout of enzyme function. ABPs use a range of chemical scaffolds, including protein-reactive natural products, mechanism-based inhibitors, and general electrophiles [32,33]. Currently, probes exist for more than a dozen enzyme classes, including proteases, phosphatases, glycosidases, oxidoreductases, and ATP-binding proteins [32,33].

ABPs often contain small chemical handles (alkyne or azide) that accept reporter tags, such as biotin and fluorophores, via click chemistry. Attachment of fluorophores allows visual detection of ABP-labeled proteins by in-gel fluorescence imaging. Biotin or other affinity tags can be used to enrich and isolate labeled proteins for identification by mass spectrometry [32,33]. ABPP can validate in silico function predictions and assign biochemical function to hypothetical proteins, providing a high throughput method to experimentally annotate the *Mtb* genome.

The power of ABPP goes beyond experimental annotation. ABBP allows for comparative analysis of proteomes in parallel. The functional proteome of different life cycle stages, *Mtb* under host-induced stress, and drug treatment can be determined to assess the dynamics of enzyme activities, and identify condition-specific targets for the treatment of TB. In addition, competitive ABPP can identify the targets and off-targets of small molecule inhibitors with unknown mechanisms of action. In a competitive ABPP assay, drug competes with ABP for protein binding and reduction in protein labeling between inhibitor-treated and untreated proteomes identifies inhibitor targets. The many applications of ABPP enable high-throughput

functional classification of the *Mtb* proteome in the context of infection and treatment-related conditions, providing a better understanding of *Mtb* physiology and pathogenesis.

ATP-binding Proteins

ATP-binding proteins are the largest group of *Mtb* proteins and drive a variety of cellular processes and molecular signaling events based on their ATP-binding and hydrolyzing capabilities. ATP-binding proteins include synthases, kinases, chaperones, and transporters, reflecting the integral nature of ATP-binding and its importance in a wide range of physiological processes that are critical to *Mtb* viability [34,35]. In addition, ATP-binding proteins play a critical role in *Mtb* virulence. For example, ATP driven efflux pumps work to reduce intracellular drug concentrations and are responsible for infection-induced drug tolerance [36,37].

ATP-binding proteins generally contain conserved sequences, such as the Walker-A and Walker-B motifs that coordinate ATP and co-factor binding important for catalysis. In addition, many ATP-binding proteins have common structural folds, such as the Rossman fold. These conserved, common elements aid in the classification of ATP-binding proteins and allow for functional prediction using in silico approaches. Sequence variation does exist and ATP-binding occurs even in the absence of canonical motifs and folds associated ATP-binding proteins, limiting the use of in silico approaches to identify and classify ATP-binding proteins. The importance of ATP-dependent enzyme functions in *Mtb* viability and virulence, and the availability of ATP analog –based drugs, make annotation of this class of proteins particularly relevant for guiding the discovery of new therapeutic targets to treat TB.

Eukaryotic-like Serine/Threonine Kinases

Among the signaling systems of *Mtb*, reversible phosphorylation on protein Ser/Thr provides a central mechanism for sensing and responding to external cues. While the classical two-component systems are underrepresented in *Mtb* compared to other bacteria, the

“eukaryotic-like” serine/threonine protein kinase (STPK) family is expanded, with 11 members [38]. The number of STPKs in prokaryotes usually correlates with more complex lifestyles that include different physiological states such as sporulation or bacteriostasis [39].

The extracellular domains of the *Mtb* STPKs provide a large and diverse repertoire of sensors to detect physiological changes in the host [40]. Of the 11 STPKs expressed in *Mtb*, only three are essential in standard *in vitro* broth culture, suggesting functions in physiological states distinct from *in vitro* replication. To date, relatively little is known about the physiological roles of the *Mtb* STPKs, but recent data implicate bacterial Ser/Thr phosphorylation in development, stress response, and host-pathogen interactions [41], processes that are also at the center of *Mtb* adaptations underlying latency and reactivation. For example, the *B. subtilis* STPK PrkC triggers exit from dormancy and the germination of spores by sensing peptidoglycan through an extracellular penicillin and Ser/Thr kinase associated (PASTA) domain [42]. In striking homology to PrkC, the essential STPK of *Mtb*, PknB, also codes for a PASTA domain [43] and could thus sense and transmit growth cues. Such a role in growth and division has indeed been shown for PknA and PknB orthologs in *M. smegmatis* and *M. bovis BCG*, although not in the context of growth limiting adaptations [44]. Overexpression of PknA and PknB both led to severe cell shape changes, producing elongated, widened, and branched cells. These phenotypes were accompanied by severely reduced growth and survival of bacteria in culture. The identification of a homologue of the cell division/cell shape regulator DivIVA, Wag31, as an *in vivo* substrate of PknA and PknB further supports the idea that PknA and PknB regulate cell growth and division [45]. Indirect evidence also links PknB to stress response through the sigma factor SigH, which is part of the Enduring Hypoxic Response and coordinates the response to several stresses [46].

Other functional roles of the *Mtb* STPKs are gradually emerging. Some STPK knockouts have been generated and promote a range of phenotypes, but most are very poorly studied and analysis under growth-limiting conditions is lacking [41,47]. A recent phosphoproteomic study

provided a global survey of *Mtb* STPK substrates. The 500 phosphorylation sites detected in 301 proteins hint at broad regulation of mycobacterial physiology by Ser/Thr phosphorylation and suggest a range of substrates for each of the eleven STPKs. Importantly, many phosphorylation events are specific to different stimuli or growth conditions such as hypoxia [48], supporting the idea that Ser/Thr phosphorylation in *Mtb* is a major signaling mechanism. Together, these data support a prominent role of Ser/Thr phosphorylation in *Mtb* adaptations, but cause and effect relationships are unknown for the vast majority of phosphorylation events.

Although the STPKs are prime candidates for controlling *Mtb* adaptations underlying latency and reactivation, no experimental evidence yet corroborated this notion. Because latency is tightly linked to intrinsic drug resistance, treatments that specifically interfere with pathways that are responsible for the latency program hold promise to target non-replicating bacterial populations. Ser/Thr kinases are established drug targets, and several human kinase inhibitors have advanced to the clinic.

Tyrosine and Dual Specificity Kinases

Bacteria possess a versatile repertoire of protein kinases that fall into distinct families, including histidine and aspartate kinases, Ser/Thr kinases, and tyrosine (Tyr) kinases. Bacterial species across all phyla support Tyr phosphorylation [49]. Although less abundant and more labile than Ser/Thr phosphorylation, Tyr phosphorylation is integral to cell processes like exopolysacchride production, virulence, and stress response [49,50]. In general, prokaryotic Tyr phosphosignaling is mediated by bacterial tyrosine (BY) kinases that do not resemble the eukaryotic Tyr kinases; however, in a recent study of Tyr phosphorylation in *E. coli,* deletion of canonical BY kinases had a limited impact on the Tyr phosphoproteome, suggesting the presence of other Tyr signaling enzymes [51].

There is a unique class of kinase with both STPK and Tyr kinase activity, the dual specificity protein kinases. Many dual specificity kinases phosphorylate in *cis* and *trans* on Ser,

Thr, and Tyr; however, a subset of dual specificity kinases autophosphorylate on Tyr but *trans* phosphorylate on Ser/Thr residues only [52-54]. Dual specificity kinases play prominent roles in eukaryotic phosphosignaling networks. In yeast, Tyr phosphorylation is exclusively carried out by dual specificity kinases. Until recently, only two bacterial species, *Chlamydophila pneumoniae* and *Bacillus anthracis*, were known to mediate Tyr phosphorylation through dual specificity kinases [55,56].

Although many organisms use reversible Tyr phosphorylation to translate extracellular cues into cellular responses, the existence of Tyr phosphorylation in *Mtb* was disputed. In silico analyses revealed that *Mtb* does not encode canonical Tyr kinases. In addition, previous phosphoproteomic screens did not detect Tyr phosphorylated proteins, although they detected hundreds of proteins phosphorylated on Ser and Thr [48]. *Mtb* encodes two protein Tyr phosphatases, *ptpA* and *ptpB* [57]. In the absence of detectable Tyr phosphorylation, PtpA and PtpB were thought to dephosphorylate host proteins. Indeed, there is evidence that PtpA dephosphorylates the host protein VPS33B to prevent lysosome maturation and *Mtb* killing within macrophages [7]. In addition to the presence of two protein Tyr phosphatases, there are two reports of *in vitro* Tyr phosphorylation – several *Mtb* Tyr phosphorylated bands detected via Western Blotting and *in vitro* Tyr kinase activity of Rv2233 [58,59]; however, we were unable to replicate and no further data corroborate these findings.

The *Mtb* phosphosignaling network is critical to *Mtb* viability and contributes to its success as a pathogen. Although Tyr phosphorylation is supported by bacterial species across many phyla, including the actinobacteria, there was no conclusive molecular evidence for protein Tyr phosphorylation in *Mtb* to date. Thus, it remained unknown whether one of the major phosphosignaling pathways exists in this bacterial pathogen.

Dissertation Summary

The results presented in this dissertation are the products of three lines of investigation towards the experimental characterization of the functional *Mtb* proteome and Ser/Thr/Tyr phosphosignaling network. Chapter 2 details a high-throughput approach to experimentally annotate the *Mtb* ATP-binding proteins. Chapter 3 describes efforts to characterize a subset of ATP-binding proteins, the Ser/Thr protein kinases, to determine their role in growth-limiting adaptations that underlie *Mtb* latency and reactivation, using the *in vitro* hypoxia system to model these disease states. Chapter 4 describes the identification of tyrosine phosphorylation and Ser/Thr protein kinases as dual specificity protein kinases in *Mtb*. The final chapter summarizes experimental findings, discusses emerging questions about ATP-binding and Ser/Thr/Tyr phosphosignaling in *Mtb*.

Chapter 2 - Novel and Widespread Adenosine Nucleotide-Binding Function in *Mycobacterium tuberculosis*

The following text is from the article: Charles Ansong, Corrie Ortega, Samuel H. Payne, Daniel H. Haft, Lacie M. Chauvigne-Hines, Michael P. Lewis, Samuel O. Purvine, Anil K. Shukla, Suereta Fortuin, Richard D. Smith, Joshua N. Adkins, Christoph Grundner, and Aaron T. Wright. (2013). Novel and Widespread Adenosine Nucleotide-Binding Function in Mycobacterium tuberculosis. **Chemistry and Biology**, 20(1): 123-33. doi: 10.1016/j.chembiol.2012.11.008. Figure numbers have been updated to conform to the formatting of this dissertation; however, the remainder of the text is as published.

Abstract

Computational prediction of protein function is frequently error-prone and incomplete. In Mycobacterium tuberculosis (*Mtb*), ~25% of all genes have no predicted function and are annotated as hypothetical proteins, severely limiting our understanding of *Mtb* pathogenicity. Here, we utilize a high-throughput quantitative activity-based protein profiling (ABPP) platform to probe, annotate, and validate ATP- binding proteins in *Mtb*. We experimentally validate prior in silico predictions of >240 proteins and identify 72 hypothetical proteins as ATP binders. ATP interacts with proteins with diverse and unrelated sequences, providing an expanded view of adenosine nucleotide binding in *Mtb*. Several hypothetical ATP binders are essential or taxonomically limited, suggesting specialized functions in mycobacterial physiology and pathogenicity.

Introduction

Determining the function of protein-coding genes in a genome remains one of the most challenging problems in the postgenomic era. For most newly sequenced bacterial genomes, 50%–70% of the protein-coding genes are assigned a function derived by inference (i.e., by

sequence similarity with previously characterized proteins), rather than by experiment, but these inferences are frequently inaccurate [60]. Additionally, 30%–50% of genes cannot be assigned function and are referred to as hypotheticals, severely limiting our ability to fundamentally understand microbial systems and to manipulate them for human benefit.

Although many excellent computational methods have been developed to predict and assign function to protein-coding genes, including homology-based [61] and genomic context-based [62] approaches, the rate at which these genes are experimentally characterized is exceedingly slow. To address this challenge, we established a chemical biology platform combining activity-based protein profiling (ABPP) with quantitative mass spectrometry-based proteomics to facilitate high-throughput experimental functional annotation. ABPP is a developing technology in functional proteomics that uses active site-directed chemical probes, termed activity-based probes (ABPs), to report directly on the functional state of enzymes within a complex biological sample [32,63]. By specifically probing protein function in a select portion of the proteome based upon shared principles of binding and reactivity, potentially all active members of a protein family can be identified simultaneously [32,63]. Herein, we apply our ABPP approach to experimentally annotate protein function across an entire protein family in the medically important pathogen *Mycobacterium tuberculosis* (*Mtb*), the causative agent of tuberculosis.

One of the largest functional classes of proteins is the ATP- binding proteins, which share ATP binding and hydrolysis as their unifying functional feature. ATP hydrolysis is a common reaction that profoundly shapes the cell's physiology. Many ATP-binding proteins can be readily identified by sequence signatures, such as the Walker A and Walker B motifs, or structurally by common folds, such as the Rossman fold. However, more divergent ATP-binding proteins are difficult to identify through sequence-based annotation, and many members of the class are likely still unknown. ATP-dependent proteins, including chaperones, kinases, and transporters, play essential roles in *Mtb* viability, infection, pathogenesis, and drug resistance

[34,35]. The importance of ATP-dependent enzyme functions and pathways make annotation of this class of proteins particularly relevant for guiding the discovery of new therapeutic targets to treat tuberculosis.

To improve the quality of the current *Mtb* genome annotation by experimental validation of in silico protein function assignments, and to discover protein function not detectable by sequence-based methods, we combined activity-based protein profiling (ABPP) and quantitative liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based proteomics to establish an experimental annotation platform: accurate mass and time AMT) tag-ABPP. We apply this technology to the broad assignment of function to the ATP-binding protein guild in *Mtb*. We identify a total of 317 ATP-binding proteins. For >70% of these proteins, our data provide experimental validation of prior in silico prediction. Importantly, we also identify a large number of proteins previously annotated as hypothetical proteins. These represent several previously unrecognized ATP-binding proteins and highlight the diversity of ATP-binding sequences in *Mtb* and other bacterial species. Our survey of the ATP binding space in *Mtb* experimentally refines the functional annotation of the *Mtb* genome and provides leads to unexpected ATP-binding protein function in *Mtb* and other bacteria. As many of the identified hypothetical proteins are both unique to Mycobacteria and essential for *in vitro* growth or infection, they reveal divergent ATP-dependent functional proteins that could serve as therapeutic targets for the treatment of tuberculosis.

Results

Probe Design and Synthesis

ATP-binding proteins constitute a large and centrally important protein guild in all organisms. Previously, a nucleotide acyl phosphate probe was developed for the labeling and characterization of ATP-binding proteins in eukaryotic proteomes by coupling ATP to biotin through a mixed anhydride on the terminal phosphate group of ATP [64,65]. This probe binds to

functional ATP-binding sites and facilitates covalent labeling through a reaction between the ε-amino groups of lysine residues and the mixed carboxylic phosphoric anhydride moiety of the probe to form a stable acetamide (Figure 1A) [66,67]. A unique advantage of this probe is that labeling is inherently linked to the hydrolysis of the ATP analog. Thus, labeling by the probe is direct evidence of phosphate hydrolysis. Although specifically designed for the labeling of kinases and ATPases, the probe was found to broadly label other ATP-binding proteins [64,65]. Other probe targets include nucleotide-binding proteins, CoA-binding proteins, and phosphate hydrolase/transfer enzymes. Reaction with the probe requires the presence of a nucleophilic amino acid residue. To minimize steric interference and improve binding, we removed the bulky biotin group from the terminal phosphate of ATP and replaced it with a click chemistry-compatible alkyne moiety giving ATP-ABP (Figure 1A) [68]. The alkyne group allows for the Cu(I)- catalyzed click chemistry addition of multifunctional tags for fluorescent detection, biotin tagging, and tagging for direct characterization of the probe-labeled amino acid residue(s) [69,70] (Figure 1A).

Global Quantitative Activity Profiling

To test the activity and selectivity of ATP-ABP, we labeled native *Mtb* proteome with ATP-ABP, appended a fluorescent Cy5.5 dye by click-chemistry, separated samples by SDS-PAGE, and visualized fluorescence of labeled proteins (Figure 1B). In the context of the *Mtb* proteome, the ATP-ABP showed labeling of distinct bands in the ABP-treated but not the untreated control sample. The nonhydrolyzable ATP analog ATPγS competed with probe labeling in a concentration-dependent manner, completely blocking probe labeling at concentrations above 1 mM. Similarly, ATP competed with probe labeling, requiring 10 mM ATP for complete blocking of probe binding. The 10- fold higher ATP concentration required for inhibition is likely due to hydrolysis of ATP, but not ATPγS, during the labeling reaction, effectively reducing the ATP concentration in the competitive inhibition study. To test the selectivity of ATP-ABP, we also tested the effect of dATP and another nucleotide, GTP, on

probe binding. Even at concentrations that lead to complete probe inhibition with ATP, dATP and GTP did not affect ATP- ABP labeling, showing that ATP-ABP is selective for ATP- binding proteins (Figure 1B) in the *Mtb* proteome.

To identify probe-labeled proteins by mass spectrometry, lysates from exponentially growing *Mtb* were labeled with ATP-ABP, followed by covalent attachment of biotin by click-chemistry and enrichment of labeled proteins on streptavidin agarose resin. Resin-bound proteins were washed to remove non-probe-labeled proteins and were digested with trypsin. Peptides were analyzed by high-resolution LC-MS/MS, and quantitative analyses were performed using the AMT tag approach as described in the Experimental Procedures [71]. Our AMT-ABPP (accurate mass and time tag- activity-based protein profiling) platform provides several advantages over conventional MudPIT approaches. These include quantitation directly from peptide signal intensities, accurate and statistically rigorous discrimination between true and false hits as described below, and no need for isotopic labeling procedures. Additionally, the AMT-tag approach is more sensitive because of utilization of LC-MS features, which alleviate the undersampling problem encountered by traditional ‘‘shotgun’’ proteomics, allowing for deeper proteome coverage [71].

To control for nonspecific probe binding, we analyzed DMSO- treated *Mtb* samples and *Mtb* samples treated with ATPγS prior to labeling. ATPγS controlled for adenosine-independent binding of the probe, whereas the DMSO-treated sample controlled for general nonspecific binding during streptavidin-based enrichment. LC-MS analyses of six probe-labeled sample replicates (ATP-ABP treated), four no-probe control sample replicates (DMSO-treated), and two ATPγS -pretreated control sample replicates (ATPγS -treated) identified a total of 794 proteins for which at least two unique peptides were measured per protein. We set the following criteria for inclusion in our further analysis: (1) a significant difference across the probe-labeled sample and the two negative control conditions as judged by ANOVA ($p < 0.05$), (2) a R5-fold higher abundance in the probe-labeled sample relative to the control samples, and (3) reproducibility of

peptide measurements across probe- labeled sample replicates. In typical quantitative proteomics analyses, an R2-fold change in abundance ($p < 0.05$) is a generally accepted threshold for significant difference. Here, we apply a stringent threshold of R5-fold change in abundance between controls and probe-labeled samples, thereby reducing false discoveries and increasing our data confidence. Using these criteria, 317 proteins were identified for further analysis, as shown in Table S1 (available online). This group of proteins represents a high-confidence set of ATP binding proteins. Because the 5-fold cutoff is stringent compared to similar studies, we also assembled a second group of 277 hits with 2- to 5-fold enrichment of probe-labeled versus control (Table S2). This group also contains many known ATP binding proteins, suggesting that this group, although with lower statistical confidence, contains true ATP binding proteins. Figure 1C shows a heatmap representation of the quantitative functional probe- labeling profile of the 317 *Mtb* proteins with >5-fold enrichment expressed as *Z* scores. The heatmap clearly shows high reproducibility within the six probe-labeled sample replicates (ATP- ABP treated; R2 = 0.89 ± 0.05), within the four no-probe control sample replicates (DMSO-treated; R2 = 0.87 ± 0.01), and within the two ATPγS-pretreated control sample replicates (ATPγS- treated; R2 = 0.98). Competition with ATPγS shows that all binding events are dependent on adenosine.

Assessing Performance of the AMT-ABPP Approach Using the ATP-ABP

Based on the proposed chemistry of probe-target interaction and prior studies on nucleotide acyl phosphate probes [64,65,68], the ATP-ABP is expected to be reactive toward a select class of proteins, including ATP-phosphohydrolases (ATPases/kinases), nucleotide (adenine, adenosine, NAD, and/or FAD) and DNA/ RNA binding proteins, acyl-phosphate reactive proteins, and acyl-CoA binding proteins. This a priori knowledge provides an opportunity to assess specificity of our AMT-ABPP approach using the ATP-ABP. To determine ATP-ABP target specificity, we surveyed the hits for functional characteristics according to

existing annotation. Classification of labeled proteins, as shown in Figure 2 and Table S1, was cross-validated by comparison of hits to proteins annotated as ATP-binders in PATRIC [72] and TBDB [73], by literature text mining, and by bioinformatics analysis using hidden Markov models (HMM) of protein families (TIGRFAM and PFAM) [74]. ATP-ABP labeling was observed among previously annotated ATPases, kinases, nucleotide binders, and acyl phosphate-reactive proteins. Of the proteins labeled, 68 proteins (20%) are annotated as ATP-phosphohydrolases, including several well-known ATP-interacting proteins, such as kinases, ATP-dependent proteases, and ATP-binding cassette (ABC) transporters. These proteins bind the ATP moiety of the probe and react directly with the mixed anhydride. The assignments of 48 of 68 proteins were confirmed by alignment with the PATRIC database's ATPase/ATP-dependent category [72]. Assignment of the remaining 20 were supported by HMM analysis and/or the literature [75]. A large portion of the labeled proteome consisted of nucleotide (adenine, adenosine, NAD, and/or FAD) binding proteins with a conserved adenine-binding motif capable of recognizing a structural element of the probe as well as containing a reactive amino acid capable of reacting with the mixed anhydride of ATP-ABP. This is in agreement with the previously reported reactivity profile of nucleotide acyl phosphate probes [64,65].

We also detected ATP-ABP binding of a large group of DNA and RNA binding proteins. To further define the ATP-ABP binding activity within this group, we compared the number of DNA to RNA binding proteins. Only 9 of the 51 proteins in this group are annotated as DNA binding proteins. Although some DNA and RNA binding proteins, such as topoisomerase, bind ATP as a cofactor, all proteins in this family recognize adenosine in the context of DNA or RNA. However, consistent with the lack of competition of dATP with ATP-ABP (Figure 1B), the bias toward RNA binding proteins suggests that ATP-ABP does not recognize DNA deoxynucleotide binding sites. Thus, although their general adenosine-binding propensity likely explains the identification of RNA binding proteins, probe binding to DNA binding proteins is more likely due to binding of ATP as a cofactor independent of DNA binding.

Additionally, 25 proteins known to bind or react with acyl-CoA molecules were labeled. This reactivity is likely governed by both adenine recognition and acyl phosphate reactivity, as these proteins are often responsible for hydrolysis of acyl-CoA. The probe also labeled 19 proteins that recognize nonnucleotide phosphates, such as pyridoxal phosphate, and that hydrolyze phosphate bonds (Figure 2, ‘‘Acyl-Phosphate Reactive’’). These enzymes recognize and react with the probe directly through the phosphate and mixed anhydride moieties. Of the proteins for which annotation is available, the remaining nine labeled proteins have no known nucleotide binding capability or phosphohydrolase activity. Seventy-three proteins annotated as hypothetical were also labeled by ATP-ABP and are discussed in greater detail below. Assigning nine of 317 proteins as nonselective, we estimate a false labeling rate of 3%, which is likely due to nonselective acylation of surface lysine residues by ATP-ABP [64]. However, we note the possibility that these proteins are in complex with ATP-binding proteins. Together, our AMT-ABPP approach provides high-throughput experimental validation of functional annotation for 245 in silico annotated members of the ATP-binding protein family and provides experimental evidence to functionally annotate 72 hypothetical proteins.

Pathway Annotation of ATP-ABP-Labeled Proteins

To test which pathways are particularly tractable for study by our ATP-ABP in terms of pathway coverage or representation of key components, the gene locus tags for ATP-ABP-labeled proteins were uploaded into the gene cluster algorithm within the TB Data- base integrated platform (http://www.TBDB.org). This software clusters genes into broad functional (pathway) categories. Not surprisingly, proteins classified as conserved hypothetical or unknown represented one of the largest classes observed, ranking second in number (Figure 3). Proteins involved in lipid metabolism, intermediary metabolism, and respiration represented the largest functional clusters. Additionally, a significant number of ATP-ABP-labeled proteins were assigned to the functional category, ‘‘Virulence, Detoxification, Adaptation,’’ in line with ATP-dependent enzymes playing essential roles in *Mtb* viability, infection, pathogenesis, and drug

resistance [34,35]. To assess any biases toward a particular functional category (i.e., overrepresentation of a particular functional category), we compared our experimentally observed classification to that from genome annotation (Table 1).

We define overrepresentation as a 2-fold greater representation of a functional category by probe-labeled compared to annotated proteins. Using this criterion, we observed the ‘‘Transcription and Translation’’ functional category to be over-represented. This is not unexpected because many of the proteins in this category play roles in cellular processes requiring interactions with adenosine nucleotide-like molecules. More interesting, our results also suggest the ‘‘Lipid metabolism’’ functional category to be overrepresented under the experimental conditions tested in this study. The particularly good coverage of this pathway is a direct result of the acyl-CoA binding activity of our probe: 34 proteins in this category can be tracked by ATP-ABP, 24 of those through their acyl-CoA binding properties. Thus, the reactivity of ATP-ABP toward acyl-CoA binding proteins provides a unique advantage for the probing of lipid metabolism. Lipid metabolism is thought to play important roles in *Mtb* virulence, but the mechanism(s) remain largely unclear. By offering broad coverage of the ‘‘Lipid metabolism’’ functional category, our AMT-ABPP approach offers the opportunity to experimentally probe aspects of lipid metabolism in *Mtb* in a variety of conditions, including infection.

Assigning Function to Hypothetical Proteins

Contributing to the experimental functional annotation of hypothetical proteins may be the most significant impact of this study. Approximately 25% of the *Mtb*-coding sequences are still classified as hypothetical proteins [31]. Of the ATP- ABP-labeled proteins, 33% (72 proteins), are annotated as hypothetical. These can be further organized into groups for which our functional annotation, combined with bioinformatic approaches, provides different levels of detail and functional insight (Figure 2). Hidden Markov model searches identified some level of homology for 33 of the 72 hypothetical proteins labeled by the ATP-ABP, showing homology consistent with nucleotide binding and/or ATPase activity (Table S1). To further examine and

verify hypothetical ATP binders, we compared our hits to a recent in silico analysis of *Mtb* hypothetical genes by the Bork group [75]. For 16 of 33 ATP-ABP-labeled hypothetical proteins, our HMM analysis matches the Bork group in silico annotation [75] (Table S1). Three additional proteins missed by our HMM analysis, but in agreement with our experimental assignment, were found by the Bork annotation to have ATPase or ATP-binding function. This analysis highlights the problem of conflicting protein function predictions that results from the use of different in silico annotation methods. Experimental data from the AMT-ABPP approach provides an excellent opportunity to resolve these discrepancies by validating protein function predictions. The experimental probe-labeling data provide critical evidence to validate otherwise purely predicted computational functional assignments.

A specific example of comparing our ATP-ABP target proteins identified in our screen to hypothetical proteins annotated as ATP binders by Bork is the hypothetical protein Rv0941c. In the Bork study, Rv0941c was annotated as a protein Ser/Thr kinase by orthology using genome context approaches. In our study, Rv0941c showed consistent ATP-ABP binding 15-fold over the control. Further sequence analysis and a literature search revealed that the Rv0941c C-terminal domain is similar to bacterial anti-sigma factors, a family with protein Ser/Thr kinase activity critically involved in regulating transcription [76]. In agreement with these findings, a structure prediction using the Phyre2 [77] server predicts a similar fold for the Rv0941c C terminus to the anti-sigma factor SpoIIAB from *Bacillus subtilis*, further supporting previous bioinformatic predictions [75,78] that Rv0941c is an anti-sigma factor protein Ser/Thr kinase. In contrast to the C-terminal domain, the N-terminal domain of Rv0941c shows sequence similarity to anti-anti-sigma factors, suggesting that Rv0941c is a functional gene regulatory module that regulates alternative sigma factors.

A second group of 36 hypotheticals does not have discernible homology to known nucleotide binding domains. Interestingly, among the 36, four are predicted to be essential for optimal growth *in vitro* [3], and four are predicted to be essential in infection [4] (Table S1).

Moreover, most of these proteins are taxonomically limited: 22 are found only in Actinobacteria, nine are limited to Mycobacteria, and three proteins are found only in *M. tuberculosis* (Rv0394c, Rv0831c, Rv1507c). With the absence of any sequence motifs that link the 36 unknowns to nucleotide binding, this group provides a large and completely unexplored set of highly likely novel nucleotide binding proteins that may reveal new nucleotide binding sequences, domains, or folds.

Further analysis of hypotheticals not previously identified by in silico prediction revealed that three genes, Rv3614c-3616c, likely form an operon. All three proteins were labeled by ATP-ABP; we validated ATP-binding of one of these three proteins, Rv3614c, by recombinant expression and labeling (Figure 4A). All three are annotated as hypothetical and lack homology to any known protein domains; all three are essential for growth during infection [4]; and all three are taxonomically restricted to Mycobacteria. Recently, they have been shown to be essential for ESX-1-dependent protein secretion (type VII secretion system, or T7SS) and *Mtb* virulence [79,80]. Mycobacterial paralogs to these three genes occur next to the AAA family ATPase of T7SS, suggesting that these may be ATP-binding proteins or part of a complex that includes ATPases. Their essential function in T7SS, and lack of sequence homology, make them a particularly exciting set of priority targets for further functional characterization.

ATP-ABP labeling of T7SS-associated hypothetical proteins also occurs in proteins encoded in the eleven-gene operon Rv0282- Rv0292, one of four extended T7SS cassettes in the genome. Rv0282 is the AAA family ATPase of this T7SS cassette and was identified in our screen with 4-fold enrichment over the negative control (Table S2). The ATP-ABP label also decorates proteins Rv0283 and Rv0284, encoded by genes that follow Rv0282 in tandem with overlapping stop/start codons that strongly suggest formation of a physical complex. Additional evidence for ATP binding of the protein complex is that Rv0284 contains three ATP/GTP binding site P loop motifs.

In summary, our experimental results and computational analyses now allow for a more

confident functional classification of a large number of hypothetical proteins as adenosine nucleotide-binding proteins, providing initial clues to the function of these *Mtb* proteins.

Targeted Analysis of Hypothetical Proteins

To confirm and further explore the nucleotide binding properties of hypothetical proteins identified in our screen, we expressed Rv0036c and Rv0831c in *Escherichia coli*, labeled the expressed proteins with ATP-ABP, and identified the probe-labeled amino acid residues by LC-MS/MS analysis. Both recombinant proteins were readily labeled with ATP-ABP, and ATP and ATPγS competed with probe binding (Figure 4A). Rv0036c is part of the TIGR03084 protein family, which is part of a larger set of probable enzymes, TIGR03083. Members of these protein families are found primarily in Actinobacteria. The function of these enzymes is uncharacterized, despite sharing sequence homology with other members of the protein family. Three out of nine members of family TIGR03083 encoded in *Mtb* H37Rv were labeled. Labeling of recombinant Rv0036c by ATP-ABP and subsequent tandem mass spectrometry analysis revealed the modification to occur at lysine 118. To confirm this assignment, we expressed a K118A mutant of Rv0036c. Mutation of K118 abrogated probe labeling, confirming K118 as the labeled residue (Figure 4B). Although this lysine is chemically suitable for labeling, it is not a conserved residue as shown by a multiple sequence alignment with its paralogs. Regions toward the N terminus of this family show local sequence similarity, indicating remote homology to members of a protein family, DinB, which includes mycothiol, acillithiol, and glutathione S-transferases [81]. Adenylyltransferase or CoA-transferase activity of Rv0036c, and other members of TIGR03083, is therefore likely.

Rv0831c is annotated as a hypothetical protein of unknown function. Of particular interest, Rv0831c has no discernible domain homology and is distributed almost exclusively within Mycobacteria, with only a few other distant sequence similarities outside the genus. Labeling of Rv0831c with ATP-ABP and subsequent tandem MS analysis revealed labeling at lysine 40 (Figure 4C). A K40A mutant of Rv0831c lost probe binding ability, confirming K40 as

the reactive nucleophile (Figure 4B). Thus, Rv0036c and Rv0831c both contain reactive lysine residues, confirming their labeling by hydrolysis of the acyl phosphate moiety of ATP-ABP and suggesting more pervasive presence of reactive lysine-based ATPases with unusual and previously unrecognized sequences (Figure 4). Finally, as an additional control for the reliability of MS-based assignment of reactive residues, we tested labeling, competition with ATPγS and ATP (Figure 4A), and identification of the site of probe labeling by MS for two serine/threonine protein kinases identified in our ATP-ABP screen, Rv0014c (PknB) and Rv0931c (PknD). Labeling of PknB was found at lysine 40, and PknD was labeled at lysine 44 (Figures 4A and S1). These sites of labeling match the expected ATP-binding site in PknB and the equivalent PknD site [82].

Experimental Validation of Structural Annotation

To complement our experimental functional annotation, we performed an experimental structural annotation [83] to further improve the *Mtb* genome annotation. Using a previously described bacterial proteogenomics pipeline [84], we analyzed global proteomic measurements of *Mtb* H37Rv to identify previously unrecognized coding regions in the genome (Experimental Procedures). These data validated 50% of the predicted *Mtb* proteome at the protein level, corrected 40 translational start site errors (Table S3), and identified an additional 15 protein-coding genes (Table S4). An example of a protein-coding gene identified by our analysis is shown in Figure 5. The open reading frame (ORF) now annotated as Rv4010 is defined by three peptides mapping to the genomic region 1113888 to 1114109, where no gene had been predicted previously. Note the presence of a canonical start codon ATG upstream of peptides defining the putative translational start site. Homology analysis of the genes revealed that most of the proteins are unannotated hypotheticals. Moreover, most are short (median length of 64 aa) and not annotated outside of Mycobacteria. Potentially because of either their length or their exclusive taxonomic distribution, the annotation of these 15 genes is sporadic within *Mtb* genomes. As of the writing of this report, the National Center for Biotechnology Information

(NCBI) lists 132 *Mtb* genome projects. Some of the previously unrecognized genes identified here are annotated in the genomes of numerous *Mtb* strains (e.g., Rv4007 annotated in >60 strains), whereas some are only annotated in a few (Rv4014 annotated in <10). This lack of annotation in other *Mtb* genomes typically represents false negatives missed during the annotation process, rather than strain diversity. The 15 identified protein coding-genes, and the 40 corrected gene models have been added to the RefSeq annotation with locus IDs Rv4000-Rv4014. The data can be down- loaded directly through the RefSeq FTP site hosted by NCBI (ftp://ftp.ncbi.nih. gov/genomes/Bacteria/Mycobacterium_ tuberculosis_H37Rv_uid57777/), and results already appear in all NCBI tools, such as BLAST.

Discussion

Functional annotation of bacterial genomes has been exceedingly challenging, and simultaneous global annotation across an entire protein functional class remains largely intractable [85]. High-throughput experimental methods are needed for functional annotation to systematically characterize bacterial genomes. Computational prediction of function in bacteria is often incomplete or wrong [86], and even in highly studied model systems, such as *E. coli*, hundreds of genes remain poorly annotated or entirely hypothetical [87]. Most experimentally characterized bacterial genes are derived from a small number of representative bacteria. This limits computational analyses to characterization of functions within gene families from a small set of bacteria for which a priori knowledge already exists. This leaves large areas of bacterial functional pathways in the dark [88]. Other gene classes, such as transcription factors and transport proteins, have little conserved sequence homology, and computational approaches are unreliable for their identification [89,90]. More- over, the high GC content and dissimilarity of *Mtb* to other prokaryotes has made computational functional annotation of its genome particularly challenging [91]; as a consequence, no functional information is available for 25% of *Mtb* proteins, highlighting the experimental challenges of protein function assignment in *Mtb*.

Even using inference methods, which assign function via gene neighborhood analyses and other exhaustive informatics approaches [75], much of the *Mtb* genome remains functionally undefined. Therefore, systematic experimental methods for elucidating gene function in *Mtb* are needed, and these may lead to novel therapeutic targets and targeted mapping of the biological pathways associated with *Mtb* viability, pathogenesis, and drug resistance.

Chemical biology approaches, in particular ABPP, have been developed to address these shortcomings [33,92]. Although ABPP is emerging as a powerful approach to comprehensively identify protein function across a defined enzyme class in a proteome, this approach has not been applied to bacterial annotation. Here, we establish an ABPP-MS platform toward the functional annotation of proteins in mycobacteria and apply the strategy toward defining the *Mtb* adenosine nucleotide binding family. We chose the ATP-binding protein functional class because of its large size, central role in *Mtb* physiology and pathogenicity, the close similarity of the probe to the natural ligand, and because probe labeling is direct experimental evidence for hydrolysis by the target protein.

Our study identified 317 proteins, a majority of which were previously annotated as ATP-binding proteins, and using a less stringent cutoff, we identify another 277 ATP binding proteins (Tables S1 and S2). Although in some cases, such as the serine/threonine protein kinases, the ATPase activity of proteins is well documented, most annotation is still inferred from sequence homology. Our data validate such inferences through direct experimental measurement. In many cases, our study provides experimental data on the function of these enzymes. The ‘‘true positives’’ for which our data confirms previous annotation provide a benchmark for the selectivity and reliability of our approach. Among the targets identified here, we estimate a false positive rate of 3%, as nine of 317 proteins were labeled, but were annotated as something other than ATP binding proteins (Figure 2).

The number of potential ATP-binding protein families in Mycobacteria is anticipated to be large. When analyzing the total complement of probe-labeled proteins, many of the ATP-binding

families were represented by only one or two probe-labeled proteins. Of the 317 probe-labeled proteins, 279 did not have at least 20% sequence homology to the other probe- labeled proteins, implying the presence of numerous different ATP-binding protein families. The most common homologous domains identified in the probe-labeled proteins were the protein kinase domain, PF00069, and the ABC transporter ATP-binding protein domain, PF00005. Our data suggest that there is a great deal of unexplored ATP-binding protein space yet to be discovered. Coverage of the entire complement of ATP-binding proteins in *Mtb* by our ABPP approach is not expected. ATP-ABP only labels proteins active under the given experimental condition. In evaluating a single growth phase in a defined medium, many proteins are likely not expressed or the conditions may not render them functionally active. Indeed, this selectivity of the ABPP approach for functional enzymes facilitates innovative experimental design and functional discovery. Future efforts will profile protein functional changes under different experimental conditions specific to infection and drug resistance.

Although any chemical activity probe differing from the natural ligand has the potential for off-target labeling, our probe comes very close to the natural ligand and should minimize off-target binding. The only chemical differences between the ATP-ABP and ATP are in the extended triphosphate moiety, with ATP- ABP carrying a mixed anhydride (acylphosphate) reactive group containing the click-chemistry compatible alkyne. Beside this one modification, the adenosine of ATP is unchanged, allowing the probe to work as a faithful ATP mimetic. The reactive acyl phosphate moiety of ATP-ABP raises the possibility that surface residues react with the probe independent of adenosine binding. To identify and exclude these labeling events, we used ATPγS to test if the adenosine moiety can compete for all probe binding. Our quantitative MS approach allowed for precise determination of ATPγS binding relative to probe binding and led to confident detection of off-target binding across all targets. Binding in the absence of concomitant adenosine binding was rare and was excluded from our analysis if competition with ATPγS was below a 5-fold cutoff ($p < 0.05$).

One major finding of this study is the identification and reclassification of 72 hypothetical proteins as ATP-binding proteins. These included 36 hypothetical proteins for which subsequent HMM profile analysis identified sequence similarity to nucleotide binding domains and 36 hypothetical proteins that do not have discernible homology to known nucleotide binding domains, including eight that are essential for growth and infection. This latter set of 36 hypothetical proteins likely represents novel families of ATP-binding proteins. Recombinant expression and sequence analysis of two hypotheticals shows that they indeed label at residues consistent with ATP binding, suggesting shared reaction mechanisms with known ATP- binding proteins. Further experimental analysis will be necessary to fully confirm their role in ATP binding and hydrolysis, but our initial sample suggests that many of these are indeed functional ATP-binding proteins. Thus, ATP binding appears to be more widespread than previously thought and can be facilitated by a much larger number of proteins with highly varied and unusual sequences. Identification of these members of the ATP binding family will aid in the annotation of other bacterial genomes and provide starting points for more generally defining the possible evolutionary solutions for ATP binding of proteins.

Experimental Procedures

Probe Synthesis

See the Supplemental Experimental Procedures.

Preparation of *M. tuberculosis* H37Rv Cell Lysates

Mtb strain H37Rv was grown in 7H9 medium to an optical density of 1 measured at 600 nm. Cells were harvested by centrifugation, washed in PBS, and lysed by bead-beating. Insoluble material was pelleted by centrifugation, and the lysates were passed twice through a 0.2 mm filter for sterilization.

Probe Labeling and Sample Preparation for SDS-PAGE Analysis

Log-phase *Mtb* H37Rv cell lysates (1 mg protein) in PBS were treated with ATP-ABP (20 mM), vortexed, and incubated for 1 hr at 37_C. Following probe incubation, proteomes were treated with an azide-derivatized Cy5.5 fluorescent reporter group (75 mM), tris(2-carboxyethyl) phosphine (TCEP, 1 mM), tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine (TBTA, pre- pared in 4:1 tert-butanol:DMSO, 100 mM), and CuSO4 (1 mM). The samples were vortexed and incubated at room temperature in the dark for 1.5 hr. SDS-PAGE loading buffer (reducing) was added to the samples, heated at 85_C for 2 min, and loaded onto a 10% Tris-Glycine gel. Gels were imaged using a Protein Simple FluorchemQ system.

Probe Labeling and Sample Preparation for LC-MS Analysis

Mtb cell lysates (1 mg protein) were treated with ATP-ABP (20 mM), DMSO (no- probe control), or ATPγS (inhibition control, 1 mM). Following addition of ATPγS, ATP-ABP (20 mM) was added. All samples were incubated for 1 hr at 37_ C. Following probe incubation, proteomes were treated with biotin-azide (36 mM), TCEP (2.5 mM), TBTA (250 mM), and CuSO4 (0.50 mM). The samples were vortexed and incubated at room temperature in the dark for 1.5 hr. Probe- labeled proteins were then enriched on streptavidin resin, reduced with TCEP, and alkylated with iodoacetamide. Proteins were digested on-resin with trypsin, and the resulting peptides collected for LC-MS analysis. For full details see the Supplemental Experimental Procedures.

LC-MS Analysis of Probe-Labeled Samples Utilizing AMT Tag Approach

Proteomics data for unlabeled, probe-labeled, and inhibitor-pretreated probe- labeled samples were generated and analyzed using the accurate mass and time (AMT) tag proteomics approach [71]. See the Supple- mental Experimental Procedures for details. To identify a protein as specifically labeled by the ATP-ABP, we required the following criteria: (1) the protein

exhibits a significant difference across the probe-labeled sample and the two negative control conditions as judged by ANOVA ($p < 0.05$), and (2) the protein exhibits R5-fold more abundance in the probe-labeled sample relative to both the no-label negative control and the inhibitor negative control sample.

Expression and Purification of *Mtb* Hypothetical Proteins

The full-length Rv0036 and Rv0831c genes were amplified from genomic *Mtb* H37Rv DNA and cloned into the pET28b expression vector in-frame with the N-terminal six-histidine tag. The vector was transformed into BL21 (DE3)- CodonPlus cells, and protein expression was induced at *A*600 of 0.6 by adding 100 mM isopropyl-1-thio-b-d-galactopyranoside. Protein was expressed for 20 hr at 20_C, cells were harvested, resuspended in 20 mM Tris (pH 7.5, 150 mM NaCl), and lysed by sonication. The lysate was cleared by centrifugation and loaded on a metal-chelating affinity column. Fractions were pooled, loaded on a gel filtration column, and eluted in 20 mM Tris (pH 7.5, 150 mM NaCl).

Analysis of ATP-ABP-Labeled Hypothetical Proteins

Purified proteins (30 mg) were labeled with ATP-ABP (20 mM), vortexed, and incubated for 1 hr at 37_C on a thermal mixer with mild agitation. The protein was denatured in 8 M urea and digested with trypsin, and the peptides were analyzed by LC-MS; see the Supplemental Experimental Procedures for details.

Global Proteome Profiling for Structural Annotation

Mtb H37Rv whole-cell lysate was prepared in PBS by French press [93]. Cell lysate was separated into cytosolic, light membrane, and cell wall subcellular fractions by centrifugation. Three replicates of whole-cell lysate and subcellular fractions were processed, and the tryptic peptides were analyzed by LC-MS. MS spectra were analyzed by the bacterial proteogenomics

pipeline using default values [84]. Briefly, tandem mass spectra were searched by Inspect against a translation of the genome (NC_000962) and subsequently rescored with PepNovo and MSGF. Searches did not include any posttranslational modifications but in accord with Inspect's searching paradigm did not require tryptic specificity. Each stop-to-stop ORF was included regardless of coding potential. We concatenated decoy records by shuffling each ORF. Significant peptide/spectrum matches (PSM) were those with an E-value of e_10 or smaller, which led to a peptide level FDR of 0.1% (spectrum level FDR 0.024%).

Figures

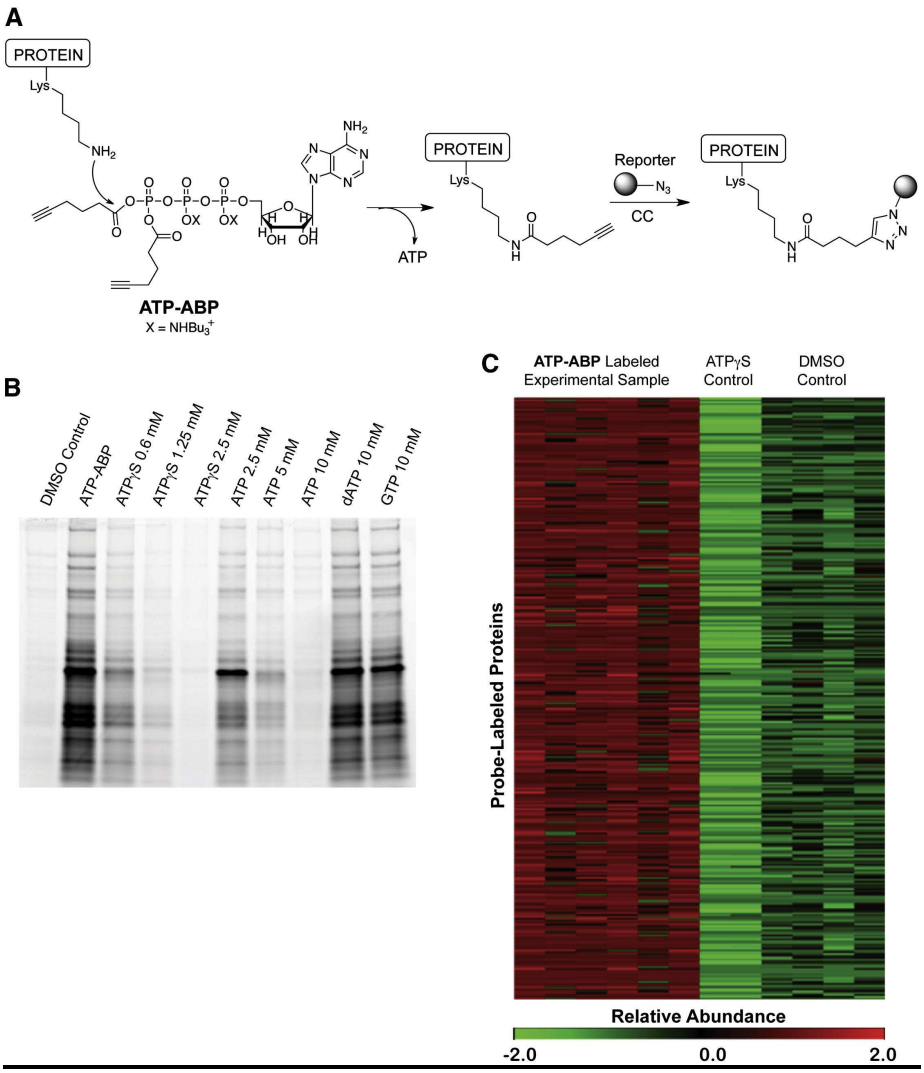


Figure 2-1. Probe structure, selective labeling, and identification of *Mtb* proteins by ATP-ABP. ATP-ABP structure and labeling of proteins. The ε-amino group of Lys residues reacts with one of two acyl phosphate moieties on the probe transferring the click-chemistry (CC) compatible alkyne unit to the protein and releasing ATP. Biotin (MS analysis) or Cy5.5 (fluorescent gel analysis) is appended to the probe-labeled protein via CC. (B) In-gel analysis of ATP-ABP-labeled *Mtb* lysate. Proteomes were labeled with ATP-ABP (20 mM) alone and in the presence of ATPγS, ATP, dATP, and GTP. Labeled proteins were visualized after SDS-PAGE. (C) Heatmap illustration of quantitative functional activity profile for 317 *Mtb* proteins, demonstrating reproducibility within probe-labeled sample replicates (ATP-ABP), no-probe control sample replicates (DMSO control) and ATPγS-pretreated control sample replicates (ATPγS control). The MS-measured protein abundances are listed in Table S1. The abundance values were converted in MultiExperiment Viewer (MeV) [94] to normalized score (z-score) for visualization. The scale is MeV normalized score (Z score) from low (green) to high (red). See also Table S2.

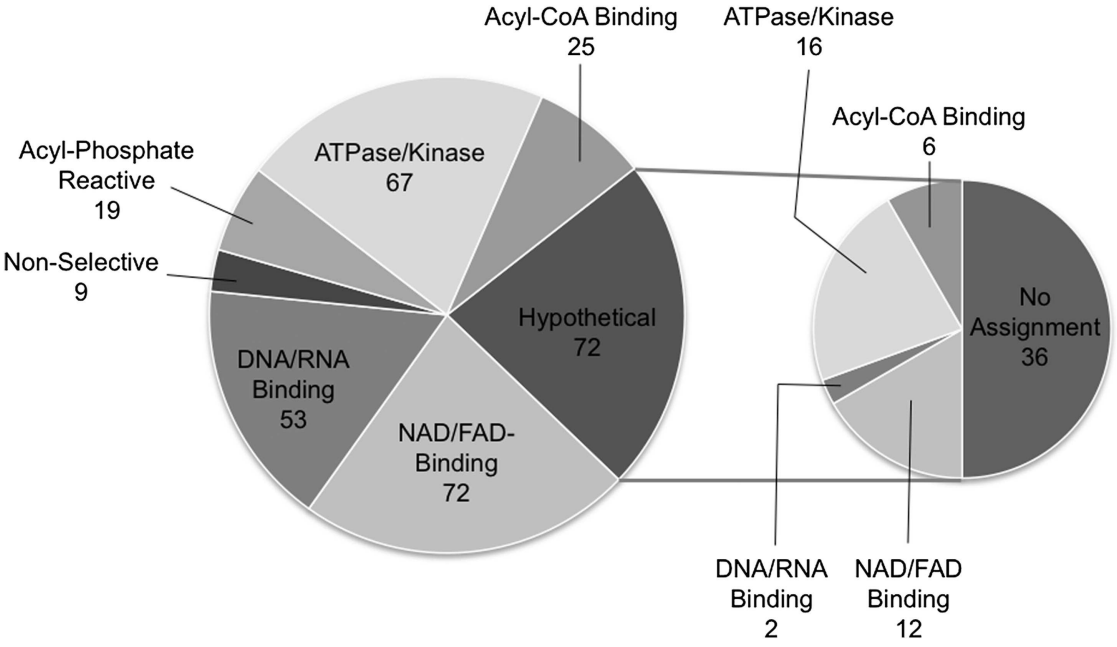


Figure 2-2. Specificity of ATP-ABP labeling in the *Mtb* proteome. Pie chart shows functional classification of 317 proteins confidently identified as interacting with the ATP-ABP based on chemistry of the ATP-ABP, literature mining, and in silico prediction. The insert pie chart shows the further functional classification of 72 hypothetical proteins identified. Approximately 45% of the hypothetical proteins were confirmed to be ATP-binding proteins by additional bioinformatics analyses. See also Tables S2, S3, and S4.

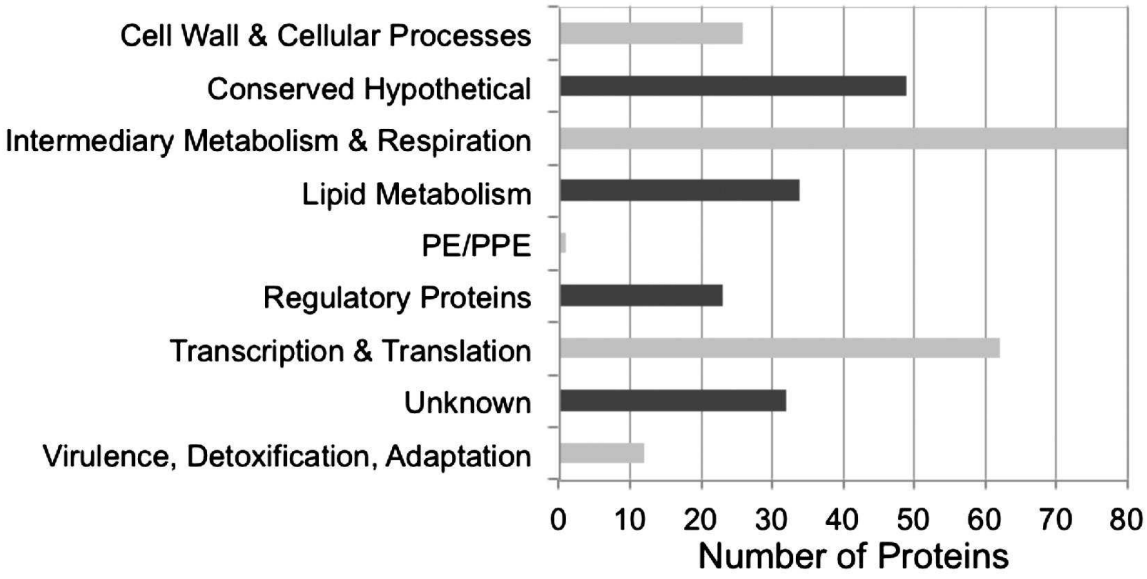


Figure 2-3. Pathway distribution of 318 ATP-ABP labeled proteins from *Mtb.* Proteins were mapped into functional processes and pathways using http:// www.TBDB.org gene mapping programs.

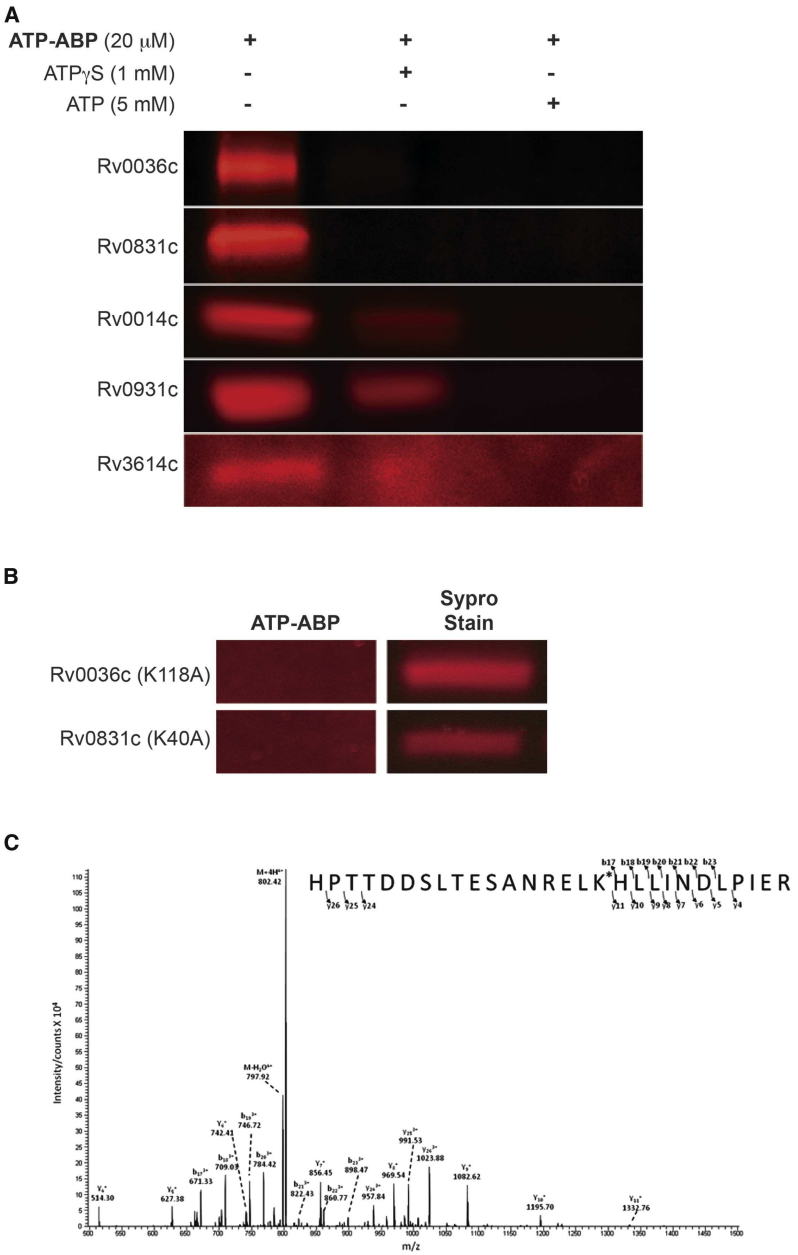


Figure 2-4. Validation of hypothetical protein labeling. (A) ATP-ABP (20 mM) labeling of recombinantly expressed hypothetical proteins Rv0036c and Rv0831c, T7SS protein Rv3614c, and Ser/Thr protein kinases Rv0014c and Rv0931c (MS spectra showing site of probe labeling are shown in Figure S1). Labeling of the hypothetical proteins was competitively inhibited by ATPγS (1 mM) and ATP (5 mM), showing adenosine-dependent probe binding. (B) ATP-ABP (20 mM) labeling of the K118A mutant of Rv0036c, and the K40A mutant of Rv0831c, revealing no probe labeling following click chemistry addition of a Cy5.5 dye and SDS-PAGE separation of proteins. Sypro Ruby Red stain indicates total protein used for labeling. (C) Annotated experimental MS/MS spectra showing labeling at lysine 40 of the peptide, ‘‘R.HPTTDSLTESANRELK*HLLINDLPIER.Q,’’ from the probe-labeled hypothetical protein RV0831c.

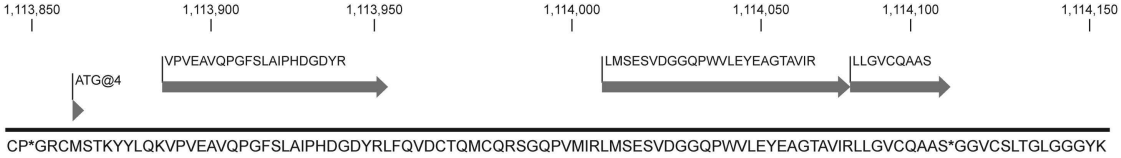


Figure 2-5. Identification of previously unrecognized *Mtb* H37Rv genes. Three peptides identified by MS-based proteomics map to the genomic region 1113888 to 1114109 on the forward strand, where no gene had been previously predicted by computational approaches. Note canonical start codon ATG upstream of peptides. See also Table S4.

Functional Categories of Observed ATP-ABP-Labeled Proteins Relative to Functional Categories Predicted from Genome

Functional Categories	Total of Genome	Percent of Genome	Total of Observed Proteome	Percent of Observed Proteome	p Value
Virulence, detoxification, adaptation	90	2.3	12	3.8	0.1221
Cell wall and cell processes	513	13	26	8.2	0.0109
Conserved hypotheticals	907	22.9	49	15.4	0.0016
Transcription and translation	206	5.2	62	19.4	0.0001
Insertion seqs and phages	136	3.4	0	0	0.0001
Intermediary metabolism and Respiration	877	22.1	80	25.1	0.2351
Lipid metabolism	225	5.7	34	10.7	0.0009
PE/PPE	165	4.2	1	0.3	0.0001
Regulatory proteins	188	4.7	23	7.2	0.0589
Unknown/unclassified	653	16.5	32	10	0.0019
Total	3,960	100	319	100	

Table 2-1. Functional categories of observed ATP-ABP-labeled proteins relative to functional categories predicted from genome. Fisher's exact test was used to calculate p values utilizing a 2 x 2 contingency table and the observation counts. Two tailed p values were calculated. Overrepresented categories are defined as ~2-fold greater ($p < 0.01$) representation of a functional category by probe-labeled compared to prediction from the genome.

Supporting Information

Supporting Experimental Procedures

Probe Labeling and Sample Preparation for LC-MS Analysis

Mtb proteomes (1 mg protein) were treated with **ATP-ABP** (20 ▪ M), DMSO (no probe control), or ATP ▪ S (inhibition control, 1 mM). Following addition of ATP ▪ S, **ATP-ABP** (20 ▪ M) was added. All samples were incubated for 1 hour at 37℃. Following probe incubation, proteomes were treated with biotin-azide (36 μM), TCEP (2.5 mM), TBTA (250 ▪ M), and CuSO4 (0.50 mM). The samples were vortexed and maintained at room temperature in the dark for 1.5 hours.

The samples were centrifuged at 11,600 × *g* for 4 minutes at 4 ℃. The supernatant was discarded, and 400 μL of ice-cold methanol was added to the pellet. The samples were sonicated on ice, vortexed, and rotated for 10 minutes at 4 ℃. The samples were centrifuged again at 11,600 × *g* for 4 minutes at 4 ℃, and the supernatant discarde d. Ice-cold methanol was added again, and the sonication, rotation, and centrifugation repeated. The supernatant was discarded. The samples were then dissolved in 520 μL of 1.2% SDS in PBS, and probe sonicated for eight one-second pulses. The samples were then heated at 95 ℃ on a thermal mixer with agitation (500 rpm). The samples were centrifuged at room temperature at 6,000 × *g* for 4 minutes. A BCA assay was performed to determine sample concentration.

For enrichment of probe-labeled proteins, a 100 μL aliquot of streptavidin agarose resin (Thermo, 1-3 mg biotinylated BSA protein per mL resin) per prepared sample was placed in a chromatography column (Biorad) on a vacuum manifold. The resin was washed 3× with PBS, and transferred to a 15 mL tube using two 0.5 mL aliquots of PBS. To the tube was added an additional 1.0 mL of PBS. To each tube was added 300 μg of protein (in 1.2% SDS in PBS). The total volume of each tube was set to 3.0 mL, giving a final SDS concentration of 0.2%. The tubes were rotated end-over-end for 4 hours at room temperature.

Following streptavidin capture of probe-labeled proteins, the solution in the tubes was transferred into the columns on the vacuum manifold, and the solution removed. The resin was then washed with 0.5% SDS in PBS (1 mL, repeat 3×), 6 M urea in PBS (1 mL, repeat 3×), MilliQ water (1 mL, repeat 3×), and PBS (1 mL, repeat 5×). The resin was then transferred to sealed 1.5 mL tube using two 0.5 mL aliquots of PBS. The tubes were centrifuged at 6,000 × *g*, and the supernatant discarded.

To reduce the disulfide bonds of the enriched proteins, urea (6.0 M) in PBS (400 µL) was added to the resin in the tubes with TCEP (5 mM), and heated at 37 ℃ on a thermal mixer for 30 minutes. Sulfhydryl groups were alkylated by addition of iodoacetamide (1 mM) with heating at 50 ℃ on a thermal mixer for 45 minutes. Following alkylation the resin was transferred back to the column and washed with PBS (1 mL, repeat 9×) followed by NH4HCO3 (25 mM, pH 8; 1 mL, repeat 5×). The resin was then transferred to a vial using two 500 µL aliquots of NH4HCO3. The tubes were then centrifuged at 6,000 × *g* for 4 minutes, and the supernatant discarded.

To obtain peptides for MS analysis, NH4HCO3 (25 mM, pH 8; 200 µL) was added to the resin for each sample along with trypsin solution (2 µL; trypsin was reconstituted in 40 µL of NH4HCO3). The resin solutions were heated at 37 ℃ with shaking for 15 hours. Following trypsin digestion, the tubes were centrifuged at 6,000 × *g* for 4 minutes, and the supernatant was collected. To the resin was added NH4HCO3 (150 µL), and the tubes were placed on a thermal mixer at 37 ℃ for 10 minutes. Again the tubes were centrifuged at 6,000 × *g* for 4 minutes, the supernatant collected, and added to the prior collection. The volatiles were then removed from the tryptic peptide solutions by speed vacuum. The dried peptides were reconstituted in NH4HCO3 (40 µL), and heated for 10 minutes at 37 ℃ with mild agitation. The samples were then centrifuged at 53,000 rpm in a Beckman TLA 120.1 rotor for 20 minutes at 4 ℃. From each sample was removed 25 µL, and vialed for subsequent MS analysis. Samples were stored at -20 ℃ until analysis.

LC-MS Analysis of Probe-Labeled Samples Utilizing AMT tag Approach

In brief, this approach utilizes tandem mass spectrometry to generate a reference peptide database (accurate mass and time tag database; AMT tag database) of observed peptides, their associated theoretical masses, and LC elution times (normalized). This database is utilized to assign peptide sequences to ion current (relative abundance) information of peptides measured using high-resolution, high mass measurement accuracy mass spectrometry (LC-MS). As part of this approach, unlabeled, probe-labeled, and inhibitor-pretreated probe-labeled samples for LC-MS analysis were analyzed using an LTQ-Orbitrap Velos™ (ThermoFisher Scientific, San Jose, CA) mass spectrometer interfaced with a reverse phase HPLC system for peptide separation (LC-MS). Peptides were reverse-phase separated on in-house manufactured columns (60 cm × 360 μm o.d. × 75 μm i.d. fused silica capillary tubing) packed with 3μm Jupiter C18 stationary phase (Phenomenex, Torrence, CA). The HPLC system was equilibrated with 100% mobile phase A (0.1% formic acid in water). Fifty minutes after peptide injection, mobile phase B (0.1% formic acid acetonitrile) displaced mobile phase A, generating an approximate exponential gradient. Split flow controlled the gradient speed operating under constant pressure (10 kpsi). Separated peptides were ionized (positive) using an electrospray ionization interface (manufactured in-house) that consisted of chemically etched electrospray emitters (150 μm o.d. 20 μm i.d) [95]. The LTQ-Orbitrap Velos™ MS was operated using a heated capillary temperature and spray voltage of 200℃ and 2.2 kV, respectively. Data was acquired for 100 minutes, beginning 65 minutes after sample injection (15 minutes into gradient). Orbitrap™ spectra were collected from 400–2000 m/z at a resolution of 100k followed by data-dependent ion trap generation of MS/MS spectra of the six most abundant ions using a collision energy of 35%. A dynamic exclusion time of 30 seconds was used to discriminate against previously analyzed ions.

Generated MS/MS spectra were searched using the SEQUEST algorithm [96] against the publicly available *Mtb* translated genome sequence (www.ncbi.gov), and re-scored using the

MS-GF approach [97]. Identified peptides of at least six amino acids in length having MS-GF score ≤1E-10, which corresponds to an estimated FDR <1% at the peptide level, were used to generate an *Mtb* AMT tag database.

Orbitrap™ spectra were deisotoped using the software tool Decon2LS [98], after which mass and elution time features were identified and matched with VIPER [99] to peptides stored in the *Mtb* AMT tag database within mass measurement accuracy and elution time accuracy cut-offs of <2 ppm and <2%, respectively. Measured arbitrary abundance for a particular peptide was determined by integrating the area under each LC–MS peak for the detected feature matching to that peptide. Matched features from each Orbitrap™ analysis (dataset) were then filtered on a false discovery rate (FDR) of less than or equal to 5%; the FDR associated with the AMT tag proteomics approach is calculated using STAC (Statistical Tools for AMT tag confidence), a statistical algorithm for assigning confidence to matched mass and elution time features [100].

Relative peptide abundance measurements in technical replicates were scaled and normalized to the data set with the least information using linear regression in DAnTE [101]. Normalized peptide abundance values were then rolled up to proteins using RRollup [101]; a minimum of five peptides was required for the Grubb's test, with a *p*-value cutoff of 0.05. Only peptides unique in identifying a single protein were utilized to estimate protein abundances. Additionally, proteins represented by <2 unique peptides were removed. Minimum values corresponding to one-half the minimal observed abundance value within data sets were imputed for missing values to allow calculations of ratios. ANOVA analyses were applied to protein abundance data sets (*p*-value ≤0.05) to identify statistically significant differences in protein expression levels.

To identify a protein as specifically labeled by the **ATP-ABP**, we required the following criteria: (i) the protein exhibits a significant difference across the probe labeled sample, and the two negative control conditions as judged by ANOVA ($p<0.05$), and (ii) the protein exhibits ≥5-

fold more abundance in the probe labeled sample relative to both the no label negative control, and the inhibitor negative control sample.

Preparation and analysis of ATP-ABP labeled recombinant proteins

Purified proteins (30 µg) were labeled with **ATP-ABP** (20 ▪ M), vortexed, and incubated for 1 hour at 37 ℃ on a thermal mixer with mild ag itation. After 1 hour, excess probe was removed by placing the sample in a 0.5 mL Amicon Ultra centrifugal filter (EMD Millipore, Billerica, MA). The sample was concentrated and washed 2× with PBS (200 µL) by repeated concentration in the filter unit to a final volume of ~30 µL. The labeled protein was denatured by addition of urea (8M), and the sample was incubated at 30 ℃ on a thermal mixer. The sample was diluted 10-fold with PBS, CaCl2 (1 mM) added, and the protein was digested with trypsin at 37 ℃ at a concentration of 1 unit trypsin per 30 u nits of protein. To remove excess salt, the samples were centrifuged (10,000 × *g*) for 2 minutes to remove any precipitate, and the supernatant was purified through a 1 mL Supelco DSC-18 solid-phase extraction (SPE) column pre-conditioned in the following manner: the SPE column was conditioned 3× with methanol (1 mL), and equilibrated 2× with 0.1% trifluoroacetic acid (TFA) in water (1 mL). The sample was passed 3× through the SPE column, the column was washed 4× with 95:5 H2O:acetonitrile (ACN) containing 0.1% TFA (1 mL), and then the bound tryptic peptides were eluted with 80:20 ACN:H2O containing 0.1% TFA (1 mL). The samples were concentrated to ~30 µL in a speedvac, and the concentration was set to 0.2 µg/µL for subsequent LC-MS analysis.

Global Proteome Profiling for Structural Annotation

Mtb H37Rv whole cell lysate was prepared in PBS by French press as described previously [93]. Additionally, whole cell lysate was separated into cytosolic, light membrane, and cell wall subcellular fractions by centrifugation [93]. Three replicates of whole cell lysate and subcellular fractions were processed. Approximately 100 ▪ g of each sample type was treated

with dithiothreitol (100 ▪ M), and incubated at 37 ℃ for 1 hour with agitatio n. To cell wall and light membrane fractions was added 1% CHAPS detergent prior to reduction. To alkylate reduced Cys residues, iodoacetamide (40 mM) was added and incubated at 37 ℃ for 1 hour with shaking in the dark. The samples were then diluted 10-fold with NH4HCO3 (50 mM, pH 8.0), and calcium chloride (1 mM) was added. To digest the proteins, trypsin was added at a 1:50 trypsin:protein ratio, and incubated for 3 hours at 37 ℃ with agitation. Following digestion, the solutions were prepped for LC-MS analysis by passing them through a C18 SPE column, or for those containing detergent, a strong cation exchange column. Sample volumes were reduced by vacuum centrifugation, and peptide concentrations were determined by BCA assay. Each subcellular fraction was further separated into 25 fractions by strong cation exchange, and each fraction was analyzed by LC-MS. MS Spectra were analyzed by the bacterial proteogenomics pipeline using default values (Venter, et al., 2011). Briefly, tandem mass spectra were searched by Inspect against a translation of the genome (NC_000962), and subsequently rescored with PepNovo and MSGF. Searches did not include any posttranslational modifications, but in accord with Inspect's searching paradigm did not require tryptic specificity. Each stop-to-stop open reading frame (ORF) was included regardless of coding potential. We concatenated decoy records by shuffling each ORF. Significant peptide/spectrum matches (PSM) were those with an E-value of e-10 or smaller, which led to a peptide level FDR of ~0.1% (spectrum level FDR ~ 0.024%).

Supplemental Figure

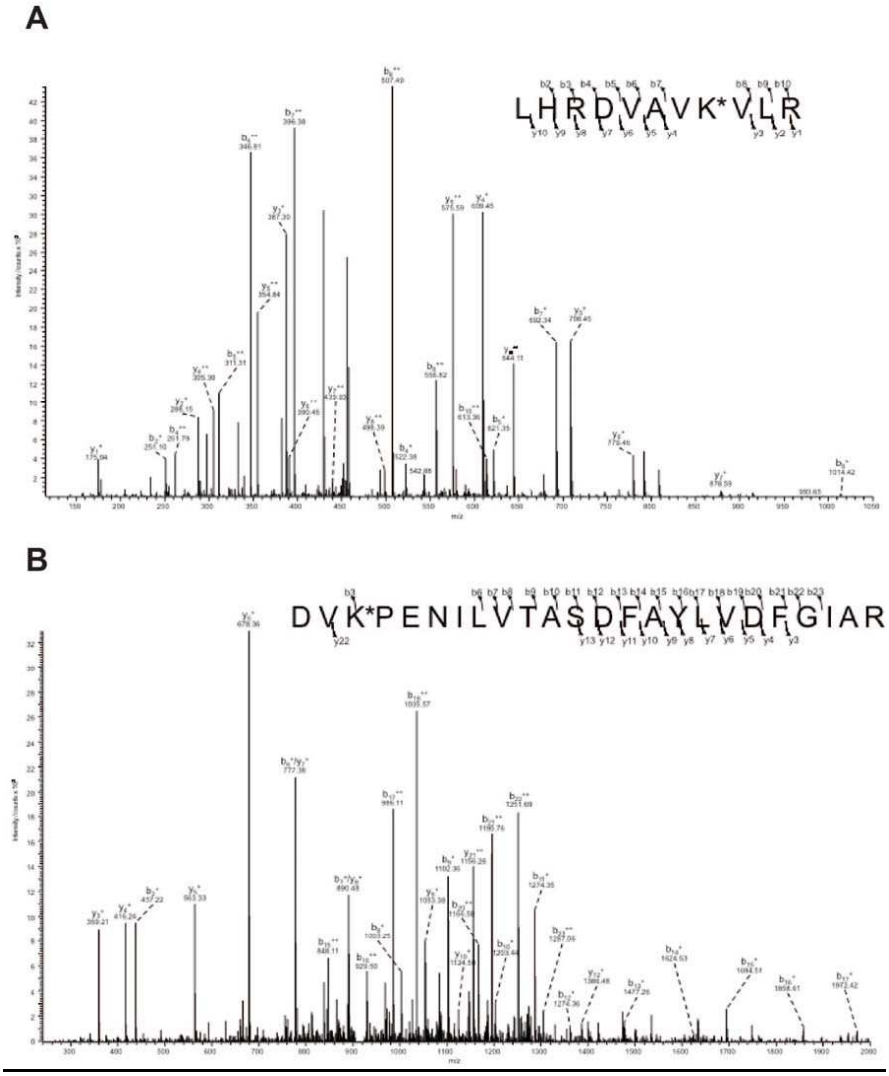


Figure 2-S1 MS spectra of the site of ATP-ABP probe labeling for (A) PknB (Rv0014c), and (B) PknD (Rv0931c). Related to Figure 4.

Supplemental Tables

http://www.sciencedirect.com/science/article/pii/S1074552112004565

Table 2-S1. Complete list of all identified proteins, separated into functional categories. MS data is shown for each replicate measurement. Hypothetical proteins show the Bork group annotation and our HMM annotation, and links proteins to prior descriptions of essentiality. The file is organized into separate data sheets. Related to Figure 1.

http://www.sciencedirect.com/science/article/pii/S1074552112004565

Table 2-S2. List of the ATP-ABP labeled proteins with a fold change from 2-5 when comparing probe to no probe control AMT values. The file is organized into separate data sheets. Related to Figures 1 & 2.

Corrected Translational Start Site Errors				
Peptide	**Genomic Location**	**Strand**	**base pairs upstream of start**	**Protein Accession**
ALEMFYDDDADLSIIQGR	NC_000962:3360542,3360595	-	9	NP_217517.1
SSANTNTSSAPDAPPR	NC_000962:4016424,4016471	+	60	NP_218091.1
SFGAHTIVKPAGPPR	NC_000962:4153776,4153820	+	84	NP_218227.2
SHSIDGGAVK	NC_000962:2131852,2131881	-	24	NP_216396.1
LDEARLELLR	NC_000962:2667203,2667232	-	117	NP_216896.1
VLDLSDGCSAGGTDMVTR	NC_000962:2113098,2113151	+	42	NP_216382.1
HSVTSSEVQTR	NC_000962:2429414,2429446	-	177	NP_216681.2
VPSFDVVFVGHR	NC_000962:57410,57445	+	at the start, but not Met	NP_214566.1
TTKDHSLATATMPNQGSSNK	NC_000962:3100143,3100202	-	33	NP_217306.1
IASPDGVAFASIDGELGEPSEMTAR	NC_000962:3350982,3351056	-	63	NP_217509.1
VIAVLMRPEPDDDWCAR	NC_000962:2123675,2123725	+	9	NP_216390.1
TDQLEDQTQGGSTVDR	NC_000962:3669407,3669454	-	66	NP_217804.2
VLHAQPPDQSTETAR	NC_000962:2698418,2698462	+	111	NP_216918.1
TTAVMSAPLAEVDPDIAELLAK	NC_000962:1220562,1220627	+	12	YP_177787.1
VAAMHYPVWR	NC_000962:3082900,3082929	+	9	NP_217291.1
LSMAEILEIFTATGQHPLK	NC_000962:4163730,4163786	+	6	NP_218237.1
TDVLPVVAVTYSPGPHLER	NC_000962:3646843,3646899	-	15	YP_177952.1
RNITTTGQIGDGR	NC_000962:1929779,1929817	-	96	NP_216219.1
FVYEDDDVVAFLTIEPMTQGHTLVVPR	NC_000962:854101,854181	-	24	NP_215273.1
LGVLDVGSNTVHLLVVDAHR	NC_000962:586352,586411	+	42	NP_215010.1
SGGACIAVR	NC_000962:4125376,4125402	+	63	NP_218201.1
HDLAQNSNSASEPDR	NC_000962:2512473,2512517	+	66	NP_216757.1
TTDVLSDTDVSLK	NC_000962:1078680,1078718	+	63	NP_215484.1
MEGDAGAGQLNPADANK	NC_000962:206748,206798	+	66	NP_214689.1
TTPTTLSGAILDPMLR	NC_000962:3649381,3649428	+	39	NP_217785.1
TTINLPNATAQLVR	NC_000962:3907002,3907043	-	36	NP_218004.1
SEAKPLHLVLGDEELLVER	NC_000962:2711260,2711316	-	15	NP_216929.1
MNDSNDTSVAGGAAGADSR	NC_000962:3031788,3031844	+	57	NP_217236.1
TDGPLIVQSDKTVLLEVDHELAGAAR	NC_000962:960092,960169	-	18	NP_215376.1
VDPTATDSPK	NC_000962:1710998,1711027	+	30	NP_216036.1
EAPDRNLAMELVR	NC_000962:1228568,1228606	-	24	NP_215615.1
AFATEHPVVAHSEYR	NC_000962:1837015,1837059	+	60	NP_216149.1
GDPPLESIVSMLSPEALTTAVDAAQQAIALADTLDVLAR	NC_000962:1858703,1858819	+	30	NP_216165.1

AAVDGKGPAAMNTHFPDAETVR	NC_000962:3496521,3496586	+	30	NP_217647.1
ASMQRPAADTPDGFGVAVVR	NC_000962:4200168,4200227	-	6	NP_218270.1
IQREPSASAHR	NC_000962:3148283,3148315	-	57	NP_217356.1
SLPEPSAALANTTTR	NC_000962:3714338,3714382	+	54	NP_217846.1
TVAAPPVCVLGLGLIGGSIMR	NC_000962:4200364,4200426	+	57	NP_218271.1
PDVLDPLGPGAPVLVAGGR	NC_000962:2416344,2416400	-	6	NP_216671.1
SRLAVDSGQVLAEPK	NC_000962:3621222,3621266	-	12	NP_217758.1

Table 2-S3. List of the 40 corrected translational start site errors. Related to Figure 2.

New Protein-Coding Genes and Corresponding Peptides

Rv4000

3 peptides, NC_000962:3392779,3393201 +

Peptide 1, LAPMWLPFR in NC_000962.Protein53572, unique=1, NC_000962:3392863,3392889 +

Peptide 2, DAHITGPYR in NC_000962.Protein53572, unique=1, NC_000962:3392986,3393012 +

Peptide 3, WWTAVGPR in NC_000962.Protein53572, unique=1, NC_000962:3393013,3393036 +

Rv4001

6 peptides, NC_000962:2510351,2510659 -

Peptide 1, GERPGFQSDSAAR in NC_000962.Protein99287, unique=1, NC_000962:2510501,2510539 -

Peptide 2, QTAPPVRPMTSDQLPATK in NC_000962.Protein99287, unique=1, NC_000962:2510447,2510500 -

Peptide 3, RPMTSDQLPATK in NC_000962.Protein99287, unique=1, NC_000962:2510447,2510482 -

Peptide 4, ADLYAAVDAMR in NC_000962.Protein99287, unique=1, NC_000962:2510414,2510446 -

Peptide 5, ADMRELLEQISTLIR in NC_000962.Protein99287, unique=1, NC_000962:2510369,2510413 -

Peptide 6, ELLEQISTLIR in NC_000962.Protein99287, unique=1, NC_000962:2510369,2510401 -

Rv4002

4 peptides, NC_000962:1242694,1242912 +

Peptide 1, AYTPDEVR in NC_000962.Protein19868, unique=1, NC_000962:1242796,1242819 +

Peptide 2, LHQRLDESDVDGYQSR in NC_000962.Protein19868, unique=1, NC_000962:1242826,1242873 +

Peptide 3, LDESDVDGYQSR in NC_000962.Protein19868, unique=1, NC_000962:1242838,1242873 +

Peptide 4, DESDVDGYQSR in NC_000962.Protein19868, unique=1, NC_000962:1242841,1242873 +

Rv4003

2 peptides, NC_000962:476310,476642 +

Peptide 1, DTDIGQPCSPEGAK in NC_000962.Protein7720, unique=1, NC_000962:476481,476522 +

Peptide 2, LWGNPGPIYCER in NC_000962.Protein7720, unique=1, NC_000962:476523,476558 +

Rv4004

2 peptides, NC_000962:665434,665616 -

Peptide 1, CKQTIEPGWLYITAHR in NC_000962.Protein128117, unique=1, NC_000962:665527,665574 -

Peptide 2, QTIEPGWLYITAHR in NC_000962.Protein128117, unique=1, NC_000962:665527,665568 -

Rv4005

8 peptides, NC_000962:2869253,2869636 -

Peptide 1, VTALESQVR in NC_000962.Protein93680, unique=1, NC_000962:2869574,2869600 -

Peptide 2, VRASEQDAAAAR in NC_000962.Protein93680, unique=1, NC_000962:2869523,2869558 -

Peptide 3, VLAGAADRDVTEFVGEFR in NC_000962.Protein93680, unique=1, NC_000962:2869469,2869522 -

Peptide 4, RATIGSFNALR in NC_000962.Protein93680, unique=1, NC_000962:2869427,2869459 -

Peptide 5, ATIGSFNALREDFTALREEMTER in NC_000962.Protein93680, unique=1, NC_000962:2869388,2869456 -

Peptide 6, FSHVEER in NC_000962.Protein93680, unique=1, NC_000962:2869367,2869387 -

Peptide 7, GKLDGAAAGQQR in NC_000962.Protein93680, unique=1, NC_000962:2869295,2869330 -

Peptide 8, IVELIEQLIADQG in NC_000962.Protein93680, unique=1, NC_000962:2869256,2869294 -

Rv4006

2 peptides, NC_000962:274710,274919 -
Peptide 1, AKHLVDIDEQALNMAR in NC_000962.Protein133841, unique=1, NC_000962:274854,274901 -
Peptide 2, TELGTTTIKDTVNAALR in NC_000962.Protein133841, unique=1, NC_000962:274803,274853 -

Rv4007

6 peptides, NC_000962:3580274,3580495 +
Peptide 1, AAVVGGGPQDEIPEADAVEQGR in NC_000962.Protein56531, unique=1, NC_000962:3580289,3580354 +
Peptide 2, AVVGGGPQDEIPEADAVEQGR in NC_000962.Protein56531, unique=1, NC_000962:3580292,3580354 +
Peptide 3, VGGGPQDEIPEADAVEQGR in NC_000962.Protein56531, unique=1, NC_000962:3580298,3580354 +
Peptide 4, GPQDEIPEADAVEQGR in NC_000962.Protein56531, unique=1, NC_000962:3580307,3580354 +
Peptide 5, AVDFDDEAGLDTAYLSGGAGDRDASEADVVDQAFVVPVADDEEIDR in NC_000962.Protein56531, unique=1, NC_000962:3580355,3580492 +
Peptide 6, DASEADVVDQAFVVPVADDEEIDR in NC_000962.Protein56531, unique=1, NC_000962:3580421,3580492 +

Rv4008

4 peptides, NC_000962:3580496,3580684 +
Peptide 1, VLTIHGVTEHGR in NC_000962.Protein56535, unique=1, NC_000962:3580577,3580612 +
Peptide 2, LTIHGVTEHGR in NC_000962.Protein56535, unique=1, NC_000962:3580580,3580612 +
Peptide 3, TIHGVTEHGR in NC_000962.Protein56535, unique=1, NC_000962:3580583,3580612 +
Peptide 4, HGVTEHGR in NC_000962.Protein56535, unique=1, NC_000962:3580589,3580612 +

Rv4009

3 peptides, NC_000962:2225841,2226131 -
Peptide 1, HALSAQLAFLESR in NC_000962.Protein103685, unique=1, NC_000962:2225973,2226011 -
Peptide 2, AGDREAQLLDILR in NC_000962.Protein103685, unique=1, NC_000962:2225934,2225972 -
Peptide 3, EQILGYDPATGV in NC_000962.Protein103685, unique=1, NC_000962:2225844,2225879 -

Rv4010

3 peptides, NC_000962:1113852,1114112 +
Peptide 1, VPVEAVQPGFSLAIPHDGDYR in NC_000962.Protein17886, unique=1, NC_000962:1113888,1113950 +
Peptide 2, LMSESVDGGQPWVLEYEAGTAVIR in NC_000962.Protein17886, unique=1, NC_000962:1114011,1114082 +
Peptide 3, LLGVCQAAS in NC_000962.Protein17886, unique=1, NC_000962:1114083,1114109 +

Rv4011

2 peptides, NC_000962:1539015,1539440 +
Peptide 1, RITEFATPQEAMEHR in NC_000962.Protein24719, unique=1, NC_000962:1539258,1539302 +
Peptide 2, SLGTLKQTHSR in NC_000962.Protein24719, unique=1, NC_000962:1539363,1539395 +

Rv4012

2 peptides, NC_000962:1282021,1282218 +
Peptide 1, GESKSPQESSSEGETK in NC_000962.Protein20495, unique=1, NC_000962:1282033,1282080 +
Peptide 2, SPQESSSEGETKR in NC_000962.Protein20495, unique=1, NC_000962:1282045,1282083 +

Rv4013
2 peptides, NC_000962:2321048,2321467 +
Peptide 1, SRTPLRPPVAPSEGVAADSVAVCR in NC_000962.Protein37014, unique=1, NC_000962:2321219,2321290 +
Peptide 2, TPLRPPVAPSEGVAADSVAVCR in NC_000962.Protein37014, unique=1, NC_000962:2321225,2321290 +
Rv4014
3 peptides, NC_000962:218390,218584 -
Peptide 1, TNYEAGTLLTCSHEGCGCR in NC_000962.Protein134688, unique=1, NC_000962:218477,218533 -
Peptide 2, VRIEVPCHCAGAGDAYR in NC_000962.Protein134688, unique=1, NC_000962:218426,218476 -
Peptide 3, IEVPCHCAGAGDAYR in NC_000962.Protein134688, unique=1, NC_000962:218426,218470 -

Table 2-S4. List of the 15 new protein-coding genes, and their representative peptides. Related to Figures 2 & 5.

Chapter 3 - *Mycobacterium tuberculosis* Ser/Thr Protein Kinase B Mediates an Oxygen-Dependent Replication Switch

The following text is from the article: Corrie Ortega, Reiling Liao, Lindsey N. Anderson, Tige Rustad, Anja R. Ollodart, Aaron T. Wright, David R. Sherman, and Christoph Grundner. (2014). *Mycobacterium tuberculosis* Ser/Thr Protein Kinase B Mediates an Oxygen-Dependent Replication Switch. **PLoS Biology** 12(1): e1001746. Figure numbers have been updated to conform to the formatting of this dissertation; however, the remainder of the text is as published.

Abstract

The majority of *Mycobacterium tuberculosis* (*Mtb*) infections are clinically latent, characterized by drug tolerance and little or no bacterial replication. Low oxygen tension is a major host factor inducing bacteriostasis, but the molecular mechanisms driving oxygen-dependent replication are poorly understood. Here, we tested the role of serine/threonine phosphorylation in the *Mtb* response to altered oxygen status, using an *in vitro* model of latency (hypoxia) and reactivation (reaeration). Broad kinase inhibition compromised survival of *Mtb* in reaeration. Activity-based protein profiling and genetic mutation identified PknB as the kinase critical for surviving hypoxia. *Mtb* replication was highly sensitive to changes in PknB levels in aerated culture, and even more so in hypoxia. A mutant overexpressing PknB specifically in hypoxia showed a 10-fold loss in viability and gross morphological defects in low oxygen conditions. In contrast, chemically reducing PknB activity during hypoxia specifically compromised resumption of growth during reaeration. These data support a model in which PknB activity is reduced to achieve bacteriostasis, and elevated when replication resumes. Together, these data show that phosphosignaling controls replicative transitions associated with latency and reactivation, that PknB is a major regulator of these transitions, and that PknB could provide a highly vulnerable therapeutic target at every step of the *Mtb* life cycle—active disease, latency, and reactivation.

Author Summary

Exposure to *Mycobacterium tuberculosis* (*Mtb*) can result in a latent form of tuberculosis (TB) infection, which can then reactivate and progress to active disease. With 1.8 billion infected persons and no tools to predict who will proceed to active disease, latency and reactivation are among the major challenges of TB treatment and control. Oxygen is one of the environmental triggers affecting the balance between latency and reactivation. Normal oxygen levels promote exponential bacterial growth in culture, but low oxygen levels (hypoxia) inhibit such growth, inducing reversible bacteriostasis. How *Mtb* regulates this oxygen-dependent replication switch, however, is still unknown. Here we tested the role of serine/threonine protein kinases—signaling molecules that transmit environmental cues into cellular responses—in this process. We found that kinase inhibition led to a bacterial survival defect and we specifically identified the PknB kinase as a critical regulator of the oxygen-dependent replication switch. *Mtb* growth was sensitive to elevated levels of PknB and this sensitivity increased in hypoxia. Inhibition of PknB activity led to defects in *Mtb* replication in reaeration. These data show that signaling through PknB modulates the growth and replication state of *Mtb* in response to oxygen, suggesting that PknB could be a drug target by which to control both the active replicating and latent non-replicating forms of *Mtb*.

Introduction

Mtb can survive for decades in an asymptomatic, clinically latent state before reactivating to active disease. Latent *Mtb* infection is estimated to affect 30% of the world's population, providing a large reservoir for reactivation to active disease [12]. Reactivation occurs in 2%–10% of latently infected individuals during their lifetimes [16,17], and in individuals with HIV co-infection, the rate of reactivation rises >20-fold [102].

The environmental factors that drive latency transitions and the *Mtb* signals associated with latency and reactivation are poorly characterized. Of these, oxygen tension may be the

best understood and is likely to be a major determinant [103]. Oxygen levels are closely linked to mycobacterial growth rate *in vitro* and *in vivo* [25,104,105]. Oxygen tension in granulomas is low [22,106]; conversely, reactivation from latency occurs mostly in the most oxygen-rich sites of the lung, the upper lobes [21]. Together, these data indicate that hypoxia shapes latency and reactivation during human infection. Latency also has a profound impact on the treatment of TB and the emergence of drug resistance. Several studies suggest that the environment in granulomas promotes bacteriostasis and phenotypic drug resistance [11,107-109].

Reversible protein Ser/Thr phosphorylation is a central mechanism for sensing and responding to external cues. While the classical two-component systems are underrepresented in *Mtb* compared to other bacteria with similar genome size, the “eukaryotic like” serine/threonine protein kinase (STPK) family is expanded, with 11 members [38,102]. Recent studies implicate bacterial STPKs in development, stress response, and host–pathogen interactions [41], processes that are also at the center of *Mtb* latency and reactivation. A role in cell wall generation and growth has been suggested for PknA and PknB in *M. smegmatis* and *M. bovis BCG* [44], and for PknB in *Mtb* [110]. A recent study provided a global survey of *Mtb* phosphoproteins [48] and identified hundreds of Ser/Thr phosphorylation sites, indicating broad regulation of mycobacterial physiology by Ser/Thr phosphorylation. Importantly, many phosphorylation events are specific to different growth conditions encountered during infection such as low pH, nitric oxide exposure, and hypoxia, further supporting the idea that Ser/Thr phosphorylation is a major signaling mechanism in response to changing environments.

Here, we used a defined *in vitro* hypoxia model to test the role of phosphosignaling in the replicative states that together define the *Mtb* life cycle: active disease (aerated growth), latency (hypoxia), and reactivation (reaeration). We identified PknB as a major regulator of the oxygen-dependent replication switch. PknB levels were reduced in hypoxia and restored in reaeration, suggesting a role for PknB in transducing growth and replication signals. Consistent with this finding, artificially elevated PknB levels during hypoxia led to killing of *Mtb*, while

reduced PknB activity affected regrowth upon reactivation. These data provide a link between phosphosignaling and oxygen-dependent replication in *Mtb*, and suggest therapeutic strategies for targeting latent and reactivating infection.

Results

STPK Inhibition Compromises Survival in Reaeration

To determine the role of STPKs in oxygen-dependent replication, we used an *in vitro* model of hypoxia and reaeration [103] in combination with chemical STPK inhibition. In this system, *Mtb* in midlog phase is placed in hypoxia (0.2% oxygen) for 7 d to induce bacteriostasis, then returned to normoxia for resumption of growth [28,29]. To test whether *Mtb* STPKs regulate oxygen-dependent replication, we treated bacteria with the broad kinase inhibitor staurosporine. Staurosporine inhibits *Mtb* growth with a minimal inhibitory concentration between 25 and 50 μM [111]. To test effects of staurosporine on oxygen-dependent replication, we first determined staurosporine concentrations that allow for unaltered aerated growth. H37Rv cultures treated with 10 μM staurosporine or less displayed growth kinetics similar to untreated *Mtb* during aerated growth (Figure 1A). We next tested the effect of 10 μM staurosporine on *Mtb* during hypoxia and reaeration. Hypoxia led to rapid growth arrest of untreated *Mtb*, and regrowth after reaeration as previously observed [24,27,103]. Similarly, staurosporine-treated *Mtb* was not compromised in hypoxia; however, staurosporine-treated *Mtb* lost viability upon reaeration in a dose-dependent manner. Concentrations as low as 0.1 μM staurosporine reduced survival as determined by colony forming unit (CFU) assay (Figure 1B) and a direct readout of ATP measuring bacterial viability (Figure S1). The 10 μM dose of staurosporine, while not affecting aerated cultures, resulted in a ~5-fold decrease in CFU and ATP levels upon reaeration. These data suggested that STPK inhibition is tolerated during hypoxia, but that one or more STPKs have a causal role in resuming growth and division of *Mtb* in reaeration.

PknF, PknD, and PknB Are the Primary Targets of Staurosporine

In mammalian cells, staurosporine is known to act on numerous STPKs [112]. This binding promiscuity limits the use of staurosporine for defining the contribution of individual STPKs to the reaeration phenotype; however, based on the binding properties of this compound in the human proteome [112], we expected that only a subset of *Mtb* STPKs binds staurosporine. To identify these targets in *Mtb*, we used a competitive activity-based protein profiling (ABPP) approach (Figure 2) [113]. We incubated detergent-solubilized *Mtb* proteome with staurosporine at 25-fold molar excess over an ATP–activity-based probe (ATP-ABP) (Figure 2A) [64,114]. Then, we labeled staurosporine-treated and -untreated samples with ATP-ABP at subsaturating concentrations. Using a click-chemistry enabled biotin tag, we purified ATP-ABP-labeled proteins from control and staurosporine-treated samples. By quantitative accurate mass and time (AMT) tag mass spectrometry [115], we quantitated the reduction in ATP-ABP binding of individual ATPases in the presence of staurosporine. ATP-ABP labeled ~200 ATPases (Figure 2B,C and Data S1) [114]. The greatest reduction of ATP-ABP binding in the presence of staurosporine was detected for three STPKs: PknB, PknD, and PknF. ATP-ABP labeling of PknF by staurosporine was strongest. No PknF peptides were detected by mass spectrometry in the staurosporine-treated samples, and the reduction of binding was arbitrarily set to 50-fold. ATP-ABP labeling of PknD was reduced 40-fold, and PknB labeling was reduced 10-fold (Figure 2C). Other staurosporine targets that showed reduced ATP-ABP labeling in the presence of staurosporine included the nonessential possible fatty acid synthase Rv3720 with almost 8-fold reduction of signal, while all other proteins showed <5-fold reduction of ATP-ABP binding (Data S1). Proteins showing less than 5-fold reduction of ATP-ABP binding are unlikely to bind staurosporine during hypoxia because the staurosporine concentration used in hypoxia was 50-fold lower than that used in the competitive ABPP screen. These ABPP data define the staurosporine binding profile in *Mtb,* identify the STPKs PknB, D, and F as its main targets, and link PknB, D, and F to the survival defects in staurosporine-treated reaerated cultures.

PknD, PknF, and PknH Are Not Required for Oxygen-Dependent Replication

To assess the individual roles of PknD and PknF in hypoxia adaptations, we obtained transposon mutants of *pknD* (*tn:pknD*) and *pknF* (*tn:pknF*) [116]. Insertions map to nucleotide positions 89 and 380 in *tn:pknD* and *tn:pknF*, respectively, truncating the N-terminal kinase domains and rendering them inactive. The *tn:pknD* and *tn:pknF* mutants grew similarly to WT in log phase (Figure S2A). To test if *pknD* and *pknF* are conditionally required in hypoxia and/or reaeration, we exposed *tn:pknD* and *tn:pknF* mutant strains to hypoxia and reaeration and assayed viability. Similar to wild type, the mutant strains ceased to replicate in hypoxia and returned to exponential growth upon reaeration with no loss in viability (Figure S2B,C). These data suggest that PknD and PknF are not required for surviving hypoxia or growth upon reaeration, although we cannot rule out the possibility that loss of STPK activity in the two single mutants is compensated for by one or more of the other STPKs. A previous study showed that PknH, which was not covered in our screen, activates the initial hypoxic response regulator DosR [117]. To test if PknH affects growth in response to hypoxia, we tested a *tn:pknH* mutant in the hypoxia time course. The *tn:pknH* mutant showed no growth defects in hypoxia or reaeration, consistent with a limited role of DosR in hypoxia [28]. Since individually PknD, PknF, and PknH are not required for hypoxic survival, we next assessed the contribution of PknB**.**

Chemical Inhibition of PknB Affects Regrowth

To determine the individual contribution of PknB, we first adopted a chemical inhibition strategy using the inhibitor K252a, which binds PknB with ~2-fold higher affinity than staurosporine [111]. We determined the highest K252a concentration that did not affect aerated *Mtb* growth to be 10 μM (Figure 3A). We then exposed an aerated culture of *Mtb* to different concentrations of K252a immediately before induction of hypoxia. Bacterial survival after 7 d of hypoxia and after 1 and 2 d of reaeration was measured by CFU assay, and by ATP levels after 2 d of reaeration (Figure 3B, Figure S3). Similar to our experiments with staurosporine, K252a showed only small effects on *Mtb* survival in hypoxia. However, K252a exposure at the start of

hypoxia caused a dose-dependent reduction in viability upon reaeration when compared to untreated *Mtb*. K252a at 10 μM reduced viability upon reaeration 5-fold on day 1 of reaeration and almost 10-fold on day 2 of reaeration (Figure 3B, Figure S3). Because we cannot rule out off-target effects of K252a, we next sought to test the role of PknB in hypoxia and reaeration by a genetic approach.

PknB Levels Are Regulated in Response to Hypoxia

Because PknB is essential [111], genetic knockout is not feasible. To identify experimental conditions for manipulating PknB levels in hypoxia, we mined microarray data [28,29] from an *Mtb* hypoxia time course to establish *pknB* transcriptional changes. *PknB* transcripts decreased ~2-fold after exposure to hypoxia and recovered to previous levels upon reaeration, and *pknB* transcript was the most down-regulated of the STPKs measured (Figure 4A). Microarray results were confirmed by qRT-PCR in an independent analysis of an identical time course [118]. To test if protein levels changed similarly to transcript levels, we probed PknB levels by Western blot using PknB-specific rabbit IgG. PknB protein levels corresponded to transcript levels, decreasing in hypoxia and increasing to aerated levels in reaeration (Figure S5A). Since Western blotting is only semiquantitative, and because the polyclonal rabbit IgG showed significant cross-reactivity producing nonspecific background, we sought to quantitate PknB levels more reliably by mass spectrometry. We measured STPK levels by quantitative mass spectrometry, using the AMT approach as described above for ABPP analysis. *Mtb* proteomes were collected from exponentially growing cultures, day 7 hypoxic cultures, and reaerated cultures and analyzed in duplicate by mass spectrometry. Consistent with our Western analysis, PknB levels were reduced on day 7 of hypoxia, but returned to aerated levels by 12 h of reaeration (Figure 4C).

PknB levels were reduced to almost below the limit of detection on hypoxia day 7, suggesting much larger reductions of PknB than observed by Western blotting, which may be attributed to the cross-reactivity of the PknB-specific rabbit IgG. Robust peptide ion abundances

were also measured for PknA, G, D, and H, but their changes were minimal compared to PknB (Figure 4A,C). The other seven STPKs not detected by mass spectrometry may be produced at low levels or may not have produced suitable peptides for detection.

These data suggest that altered PknB activity accompanies and may be causal for hypoxia-mediated transitions. These data also suggested a genetic strategy to test this hypothesis by maintaining PknB at elevated levels during hypoxia. Thus, we generated a *pknB* mutant by recombineering [119], introducing a tetracycline-inducible promoter upstream of *pknB* (Figure S4). Insertion of the tet promoter uncoupled the expression of *pknB* from the wild-type promoter. Because *pknB* is the last gene in the Rv0014c-Rv0020c operon, the promoter insertion is expected to affect only *pknB* expression. The correct insertion of the promoter cassette was confirmed by sequencing of genomic DNA. Because the tet promoter is leaky, the *tet-pknB* mutant expressed PknB at levels ~4.5-fold higher than wild type even in the absence of anhydrotetricycline (ATc), as determined by mass spectrometry. Importantly, the uninduced *tet-pknB* mutant maintained elevated levels of PknB throughout the hypoxia time course as observed by Western blotting and mass spectrometry, allowing us to test the effects of PknB dysregulation (Figure S5).

Aerated Growth Is Sensitive to Changes in PknB Abundance

To test if altered PknB levels affect *Mtb* growth, and to establish experimental conditions suitable for testing the role of PknB in hypoxia, we assessed the effect of increasing PknB levels using the *tet-pknB* mutant. Analysis of PknB protein levels by Western blot showed ATc-dependent induction of PknB expression (Figure 5A), which quickly resulted in a severe growth defect. ATc concentrations above 10 ng/ml caused growth arrest and cell killing as determined by CFU assay (Figure 5B). To test whether growth arrest was a result of PknB kinase activity, another PknB function, or an overexpression artifact, we generated a PknB kinase-dead overexpressing mutant, *tet-pknB K40A*. In contrast to *tet-pknB, tet-pknB K40A* had no effect on

aerated growth at the highest levels of ATc induction (Figure 5C). Thus, the *Mtb* growth defect is indeed due to increased PknB kinase activity.

PknB Controls Oxygen-Dependent Replication

The uninduced *tet-pknB* mutant expressed ~4.5-fold more PknB than wild type in aerated culture but demonstrated similar growth kinetics. Also, in contrast to the wild-type strain, the *tet-pknB* mutant expressed PknB in hypoxia (Figure S5). These properties of the *tet-pknB* mutant allowed testing for effects of PknB on hypoxic survival without affecting aerated growth. We exposed wild-type and uninduced *tet-pknB* mutant strains to hypoxia followed by reaeration. Hypoxia led to rapid growth arrest of wild-type *Mtb*, and regrowth after reaeration as previously observed [103]. During hypoxia, maintaining PknB expression at the level of aerated culture led to striking killing of the mutant. After 7 d of hypoxia, the number of *tet-pknB* was reduced 10-fold compared to wild-type *Mtb* (Figure 6A). To confirm that this *tet-pknB* survival defect was due to increased PknB activity in hypoxia, we treated *tet-pknB* with K252a in an attempt to chemically revert the effects of PknB overexpression. The *tet-pknB* mutant was treated with 5 μM K252a prior to hypoxia, subjected to hypoxia for 7 d, and reaerated for 4 d. K252a-treated *tet-pknB* entered bacteriostasis in hypoxia and returned to normal growth upon reaeration, similar to wild type (Figure 6B). The ability of K252a to fully revert the PknB mediated killing in hypoxia further supports that K252a targets PknB and that increased PknB activity is responsible for *Mtb* killing in hypoxia.

To test whether PknB regulates bacteriostasis in general or specifically in response to oxygen, we tested the *tet-PknB* mutant in three other conditions encountered during infection that induce bacteriostasis: nitric oxide, low pH, and nutrient starvation. Growth and viability of the *tet-pknB* mutant were similar to wild type in all three conditions (Figure S6A, B, and C), suggesting that PknB specifically regulates growth in response to oxygen.

pknB Overexpression Causes Changes in Cell Shape

To test the morphological effects of elevated PknB levels, we imaged cells by scanning surface electron microscopy in aerated and hypoxic growth. Strikingly, uninduced *tet-pknB* mutants displayed an elongated morphology compared to wild-type cells. A subset of *tet-pknB* mutant cells was as much as 2-fold longer than wild-type cells, in particular in hypoxia. To quantify these differences in cell length, we determined the approximate size of >800 cells. The length of wild-type bacilli was 2–3 μm, with 2%–3% longer cells (Figure 7A, B, and E). In contrast, 20% of *tet-pknB* mutant cells in aerated and 50% in hypoxic culture were 4–6 μm long. In addition, the average length of *tet-pknB* mutant cells increased during hypoxia (Figure 7C,D). Because overexpression of PknB in the *tet-pknB* mutant compared to wild type was only modest, and the range of PknB expression separating the phenotypes was small, higher levels of overexpression may cause more pronounced phenotypes in hypoxia. However, higher levels would also cause growth defects in aerated cultures, complicating experiments and analysis. Thus, the experimental window to test the effect of PknB in hypoxia is small, and the 10-fold reduction in CFU might underestimate the full impact of PknB on hypoxic survival.

Combined, these data establish the first link between Ser/Thr phosphosignaling and oxygen-dependent replication and identify PknB as a major regulator of this replication switch. These studies point to finely tuned regulation of growth in response to hypoxia through PknB and highlight the distinct effects of reduced and elevated PknB levels on *Mtb* growth and replication, suggesting that PknB is a highly vulnerable target for therapeutic interference with nonreplicating and reactivating *Mtb.*

Discussion

TB pathogenesis is closely linked to *Mtb*'s ability to enter a nonreplicating state, which in turn underlies latent infection and drug tolerance. Although low oxygen tension is likely to be a

major determinant of latency, little is known about the regulation of replication in response to oxygen [103]. As a consequence, only few therapeutic targets in latent, nonreplicating bacteria are known. With a genome size of ~4.4 Mb, *Mtb* codes for a relatively large number of STPKs [2,102], and recent phosphoproteomic data show that Ser/Thr phosphorylation varies under different growth conditions, including low oxygen tension [48], together suggesting that the STPKs regulate specialized *Mtb* adaptations. However, a causal link between *Mtb* phosphosignaling and the response to oxygen has not been established.

The 11 *Mtb* STPKs may have redundant function, and two are essential, complicating genetic approaches. To overcome this hurdle, we used a chemical biology approach to test for a causal role of Ser/Thr phosphosignaling in the response to oxygen. While the broad kinase inhibitor staurosporine was a useful tool to reveal a functional link between STPKs and hypoxia survival, its binding promiscuity precluded direct identification of the responsible kinase(s). However, using staurosporine in combination with competitive ABPP allowed us to identify the STPK(s) responsible for the staurosporine phenotype. This approach has several advantages over standard loss-of-function genetics. With 11 STPKs and potential redundancy of their functions, genetic approaches are limited. The use of the generic kinase inhibitor staurosporine in combination with ABPP, in contrast, allowed for broad screening and specific identification of STPKs that mediate the staurosporine phenotype. This approach identified three STPK staurosporine targets (PknB, PknD, and PknF), providing a tractable number for genetic determination of the individual contributions of STPKs to oxygen-dependent replication. In addition, these data provide a comprehensive staurosporine binding profile in the *Mtb* proteome. Although we used a broad specificity kinase inhibitor in our studies, our MS analysis shows that staurosporine bound only a subset of STPKs. Thus, although our experiments unequivocally link PknB with oxygen-dependent growth transitions, we did not sample all STPKs, and other STPKs not sensitive to staurosporine might also contribute to hypoxic adaptations.

Changing *pknB* levels in either direction showed large effects on cell growth even in aerated culture, and the experimental window to test *pknB* effects in hypoxia was small. To determine the best genetic approach for capturing the specific contribution of *pknB* to oxygen-dependent replication, we first established the transcript and protein abundance of PknB in aerated, hypoxic, and reaerated culture. Quantitative mass spectrometry showed a dramatic decrease in PknB protein levels in hypoxia. This reduction cannot be explained by the relatively modest changes in *pknB* transcript levels, but points towards specific PknB degradation. While many STPKs are regulated by reversible phosphorylation [120], PknB appears to be regulated also on the level of protein abundance, a STPK regulatory mechanism that implies the contribution of specific proteolysis in the PknB regulatory pathway. In addition, these experiments suggested a genetic approach that allows for hypoxia-specific *pknB* expression above wild-type levels specifically during hypoxia.

Our data show that PknB levels are carefully calibrated during hypoxia and reaeration, consistent with a role of PknB in replication, which is essential in reaeration but detrimental during hypoxia (Figure 8). Changes in PknB levels have dramatic effects on cell survival that are also reflected by dramatic changes in cell morphology. The highly elongated cells resulting from PknB overexpression in aerated growth and particularly in hypoxia point to a major role of PknB in cell elongation and/or division. Elevated PknB levels drive cell elongation, but elongation appears to be decoupled from cell division, particularly in hypoxia. A previous study demonstrated widened and bulging bacilli upon overexpression of *Mtb pknB* in *M. smegmatis* and *M. bovis* [44]. This difference in phenotype from the one described here may be a result of cross-species expression in the former study.

PknB specifically regulates growth in response to oxygen, as other stresses encountered *in vivo* did not affect the survival of the *tet-pknB* mutant. How does PknB sense changing oxygen tension? The four extracellular penicillin binding protein and serine/threonine kinase associated (PASTA) domains of PknB do not suggest direct oxygen sensing [43]. The PASTA

sensor domains appear to regulate localization of PknB to the cell poles and septum through binding of muropeptides [121]. We propose that PknB indirectly senses shifting oxygen tension, for example through phosphorylation by another STPK or a two-component system. Crosstalk between STPKs and between STPKs and two component systems in *Mtb* has previously been shown [44,117]. For example, the STPK PknH phosphorylates DosR, providing a possible connection between STPK and two-component system signaling in response to hypoxia. However, we did not detect an effect of this STPK on survival in hypoxia. These data likely reflect the limited role of DosR in the response to hypoxia [28]. Alternatively, an oxygen sensor might affect peptidoglycan levels or structure that is sensed through the PknB PASTA domains and provides appropriate localization of PknB. The role of PknB in adapting to changing oxygen and replication invites a comparison to the role of another bacterial STPK carrying PASTA domains, the *B. subtilis* STPK PrkC. In *B. subtilis*, PrkC controls the germination of spores in response to peptidoglycan fragments and free muropeptides [42]. The similarities between stress response programs between these two organisms suggest conservation of linked sensing of oxygen and peptidoglycan signals across different phyla. The functional consequences of PknB are ultimately defined by its substrates. By identifying the first direct link between Ser/Thr phosphosignaling and latency-related adaptations, this study provides evidence that the downstream hypoxia effectors are at least in part phosphoproteins and specific PknB substrates, suggesting new experimental approaches to identify these targets.

This study also provides a first measure of PknB's vulnerability, an important metric for drug targets. PknB is currently viewed as a viable drug target in actively growing *Mtb* [122-124]. Our data show that PknB levels are dramatically reduced during hypoxia, requiring PknB synthesis in reaeration. As a result, reduced target levels decrease the minimal inhibitory concentration for PknB inhibitors during reaeration, and by extension in reactivation. Prevention of reactivation could have significant benefit, especially in settings with high HIV burdens and the associated high rates of reactivating disease. In contrast to reactivating disease, our data

suggest that an effective strategy to kill nonreplicating *Mtb* would be through PknB activation rather than inhibition. Chemical activation of STPKs is not well understood, and common mechanisms of chemical STPK activation are not known. Yet the concept of therapeutic STPK activation is not without precedent. The lactone compound bryostatin is under active development as a protein kinase C activator [125], and bryostatin leads to PrkC activation and the exit from dormancy in *B. subtilis* [42]. Similarly, PknB activation could drive PknB activity during nonreplication, leading to growth signals at a time when growth is deadly.

Together, our data show that Ser/Thr phosphosignaling regulates replication in response to oxygen. We identify PknB as a critical regulator of oxygen-dependent replication, and by extension of latency and the progression to active disease. By reacting to oxygen changes, PknB triggers adaptations to low oxygen stress and is itself down-regulated in the effort to cease growth. Even small increases in PknB levels during nonreplication lead to bacterial killing. As growth resumes upon reactivation, PknB activity is required and loss of activity at this stage also leads to killing. In this way, PknB provides a potential, highly vulnerable drug target to target *Mtb* at every stage of its life cycle.

Experimental Procedures

Strains, Culture Conditions, and Hypoxic Model

Frozen aliquots of each strain were thawed and expanded to generate working stocks before experiments. H37Rv (ATCC 27294), *tn:pknD*, *tn:pknF*, *tn:pknH*, and *tet-pknB* were grown at 37 ℃ in Middlebrook 7H9 supplemented with OADC and 0. 05% Tween 80 in rolling culture. The hypoxia time course was carried out as previously described [28]. Briefly, cultures were grown to reach early log phase (OD_{600} 0.1–0.3) and placed in 250 ml spinner flasks to stir at 60 RPM with low oxygen gas (0.2% O_2 balanced with N_2) flowing over the culture for 7 d. For reaeration, cultures were returned to normoxia and grown for an additional 4 d to fully reaerate.

The *Mtb tet-pknB* mutant strain was created by inserting the tetracycline regulation elements *tetO* and *tetR* along with the hygromycin resistance gene 5' to the *pknB* start codon, and a FLAG tag at the *pknB* N terminus (Figure S5). *Mtb* H37Rv genomic DNA was used as the template to amplify the 500 bp regions flanking the 3' end of the *pknA* gene and the 5' end of the *pknB* gene. The PCR fragments were then cloned into vector pRPK339. Amplification of the recombineering substrate was performed by high fidelity PCR. Electroporation was performed as described previously [126]. Colonies were selected on 7H10 plates with hygromycin. Resistant colonies were grown to an OD_{600} of 1, genomic DNA prepared, and screened for correct insertion by DNA sequencing of genomic DNA.

To induce *pknB* expression in the *tet-pknB* mutant, cultures were grown to an OD_{600} of 0.1, and anhydrotetracycline was added at concentrations from 1–20 ng/ml. Growth was monitored by OD_{600} for 6 d postinduction.

Nitric Oxide, pH, and Starvation Experiments

For nitric oxide experiments, liquid cultures were grown to early log phase (OD 0.1–0.3) and then treated daily with 100 µM DETA-NO (Sigma) for 4 d [127]. On the 5th day, cultures were centrifuged at 4,000 × g for 5 min and the medium replaced with 7H9 without DETA-NO. Growth was measured daily by OD_{600} and CFU for a total of 7 d.

For pH experiments, cultures were grown to early log phase (OD 0.1–0.3) and then centrifuged at 4,000 × g for 5 min. Medium was replaced with 7H9 at pH 4.5 [128]. Growth was measured daily by OD_{600} and CFU for 5 d. For starvation experiments, 7H9 medium was replaced with 1× PBS. Growth was measured by OD_{600} and CFU for 4 wk.

RNA Isolation and Microarrays

RNA was isolated as previously described [103]. Briefly, cultures were grown to an optical density at 600 nm of 0.1–0.3. We centrifuged 25 ml of culture at 4,000 × g for 5 min and pellets

were resuspended in 1 ml of Trizol (Gibco-BRL). The sample was lysed by bead beating 3 times at 5.5 m/s for 30-s intervals, with cooling between intervals. Lysis was followed by centrifugation for 1 min at maximum speed. Supernatant was transferred to a Phase-loc tube (5 Prime) containing 300 μl of chloroform:isoamyl alchohol (1:24) and centrifuged for 5 min at maximum speed. The aqueous layer was transferred to 1.5 ml tubes containing 270 μl of isopropanol and 270 μl of high salt solution (0.8 M Na Citrate, 1.2 M NaCl) and precipitated overnight at 4 ℃. The samples were then centrifuged for 10 min at 14,000 g at 4 ℃, supernatant was decanted, and RNA purification was performed as described in the Qiagen RNeasy Purification Kit protocol. Microarray data were obtained from a previously published study [24].

Cell Lysate Preparation

Cell lysate was prepared from 100 ml of culture. Bacteria were centrifuged at 4,000 × g, and pellets resuspended in 500 μl of lysis buffer (1× PBS, 0.5% SDS, NaF, β-glycerophosphate, and protease inhibitor cocktail (Sigma)). Samples were lysed by bead beating three times at 5.5 m/s for 30-s intervals, with cooling on ice between intervals. Samples were centrifuged at 14,000 × g for 2 min, and supernatant was sterile filtered twice with 0.2 μM filters. Cell lysates were stored at –80 ℃.

Western Blotting

We separated 20 μg of log phase, hypoxic, and reaerated cell lysates by 4%–12% SDS-PAGE. Protein was transferred to nitrocellulose membrane using a semidry transfer system. Membranes were blocked with 3% BSA in 1× PBS/0.1% Tween 20 for 1 h and incubated overnight at 4 ℃ with either 1:500 of antiphosphoserine and antiphosphothreonine antibody (Millipore), 1:2,000 of rabbit anti-PknB IgG purified from rabbit serum on protein G sepharose, or 1:1,000 M2 anti-FLAG antibody. Membranes were washed in 1× PBS/0.1% Tween 20 and

incubated for 1 h at room temperature with either 1:15,000 anti-rabbit or anti-mouse fluorophore-conjugated secondary antibody (LI-COR Biosciences). Membranes were washed in 1× PBS – 0.1% Tween 20 and imaged using an Odyssey imager (LI-COR Biosciences).

Competitive ABPP

Detergent-solubilized lysate (1 mg per sample) was incubated with 500 μM staurosporine (BioSciences) for 30 min at 37 ℃. Then, treated and untreated samples were labeled with 20 μM ATP-ABP for 1 h at 37 ℃, followed by treatment with biotin-azide (36 μM), TCEP (2.5 mM), TBTA (250 μM), and $CuSO_4$ (0.50 mM). The samples were vortexed and incubated at room temperature in the dark for 1.5 h. Probe-labeled proteins were enriched on streptavidin agarose resin, reduced with TCEP, and alkylated with iodoacetamide. Finally, proteins were digested on-resin with trypsin, and the resulting peptides collected for LC-MS analysis.

LC-MS Analysis of Competitive ABPP and Global Proteomics

Quantitative proteomics data were generated and analyzed using the AMT tag approach [71]. Tryptic peptides from labeled and unlabeled samples were collected for LC-MS analysis as described previously [28] and analyzed using a LTQ-Orbitrap Velos (Thermo Fisher Scientific) MS interfaced with a reverse phase HPLC system. MS/MS spectra were searched using the SEQUEST algorithm (V27, revision 12) [96] against the publicly available H37Rv translated genome sequence, and re-scored using the MS-GF approach [100]. Peptides at least six amino acids in length with MS-GF scores ≤1E-10, which corresponds to an estimated false discovery rate <1% at the peptide level, were used to generate an AMT tag database. Mass and elution time features were identified and matched with VIPER to peptides stored in the AMT tag database within mass measurement accuracy and elution time accuracy cutoffs of <2 ppm and <2%, respectively. The measured abundance value for a particular peptide was determined by

integrating the area under each LC-MS peak for the detected feature matching to that peptide. Matched features were then filtered on a false discovery rate of ≤5% using STAC [100]. Peptide abundance measurements in technical replicates were scaled and normalized to the data set with the least information using linear regression in DAnTE [101], then rolled up to proteins using RRollup [101]. A minimum of five peptides was required for the Grubb's test, with a *p* value cutoff of 0.05. Only peptides unique in identifying a single protein were utilized to estimate protein abundances. Additionally, proteins represented by <2 unique peptides were excluded. If peptides for a given protein were not measured in at least half the replicates for a given sample type, the protein was excluded from further analysis.

Inhibitor Experiments

H37Rv cultures were grown for three doublings and diluted to an optical density of 0.05–0.1 at 600 nm before treatment with inhibitor. Growth was determined daily by optical density for 7 d. For hypoxia inhibitor experiments, cultures were grown for three doublings to reach early log phase, placed in 250 ml spinner flasks, and inhibitor was added. Cultures were immediately subjected to low oxygen gas (0.2% O_2 balanced with N_2) for 7 d, followed by normoxia for 4 d to reaerate. Samples were taken at day 0, day 7 of hypoxia, and every other day of reaeration to measure viability by CFUs and BacTiter-Glo (Promega), as described in the manufacturer's protocol.

Acknowledgements

We thank Dr. Tom Alber for providing anti-PknB rabbit serum, Dr. Christopher Sassetti for helpful discussions, the Tuberculosis Animal Research and Gene Evaluation Taskforce for providing transposon mutants, and the Fred Hutchinson Cancer Research Institute Electron Microscopy Core (Sharmon Knecht, Stephen MacFarlane, and Bobbie Schneider) for sample preparation and electron microscopy imaging.

Figures

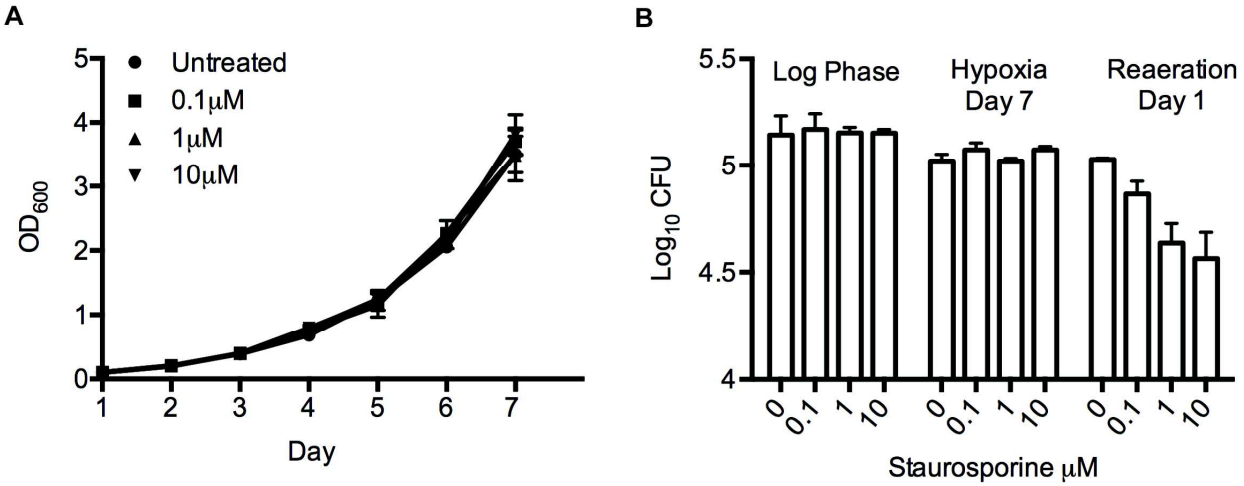


Figure 3-1. Broad STPK inhibition compromises *Mtb* growth in reaeration. (A) Wild-type *Mtb* in aerated log phase culture was treated with sub-MIC concentrations (0.1, 1, and 10 μM) of staurosporine, and growth was measured over 7 d. (B) Viability was determined on day 0, day 7 of hypoxia, and day 1 of reaeration by plating for CFU analysis. CFUs were compared by two-way ANOVA ($p < .0001$). Error bars represent standard deviation.

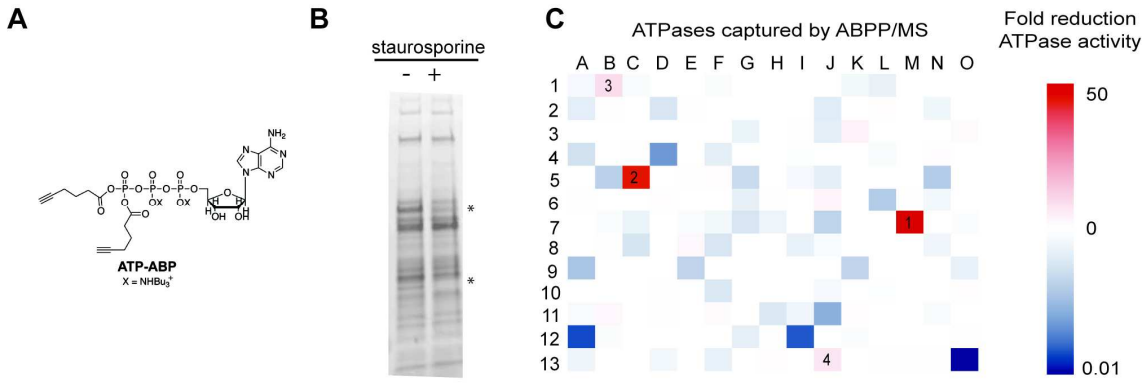


Figure 3-2. Competitive ABPP identifies PknB, D, and F as staurosporine targets. (A) Structure of the ATP activity probe. (B) Grey scale image of *Mtb* lysate fluorescently labeled following ABPP. Detergent-solubilized *Mtb* proteome was treated with staurosporine at 25-fold molar excess over ATP-ABP before labeling with ATP-ABP at subsaturating concentrations and analysis by SDS-PAGE. Probe shows labeling of ATPases and competition of staurosporine with binding to some ATPases (asterisks). (C) Heatmap representation of MS analysis of competitive ABPP to identify staurosporine targets. ATPase abundance was measured by quantitative, AMT tag approach mass spectrometry. Each square represents one *Mtb* H37Rv protein, ordered by Rv number from left to right. Color-coding represents the average fold-reduction of protein quantity in staurosporine-treated versus untreated samples from six technical replicates. Reduced ATP-ABPP probe binding in the presence of staurosporine identifies PknF, PknE, and PknB as the primary targets of staurosporine.

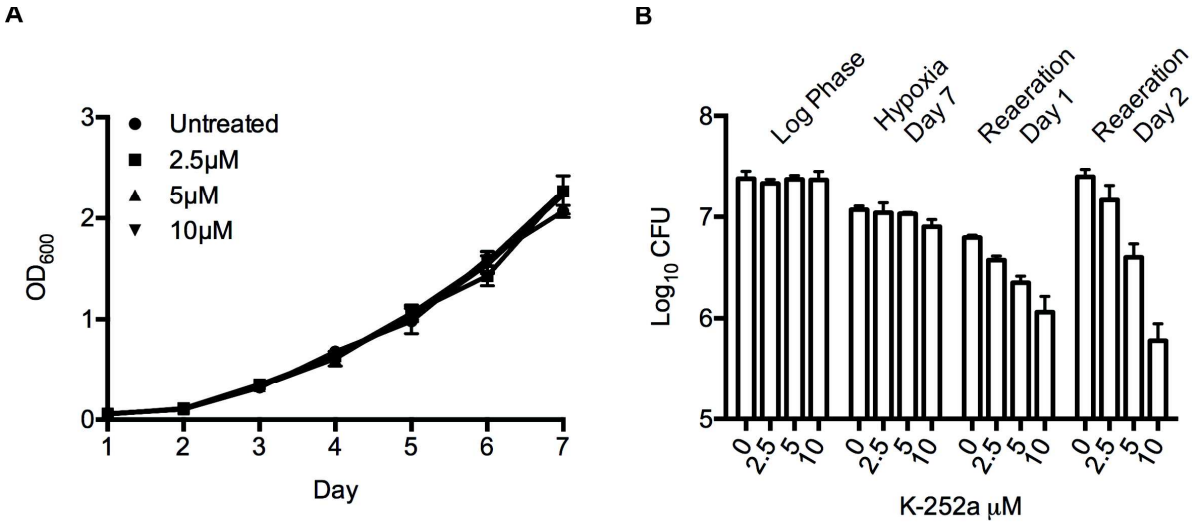


Figure 3-3. Chemical inhibition of PknB compromises viability in reaeration. (A) Wild-type *Mtb* in log phase was treated with sub-MIC concentrations of K252a, and growth in aerated culture was measured for 7 d. (B) Viability during a hypoxia time course was determined on day 0, hypoxia day 7, and reaeration day 1 and 2 by CFU analysis. CFUs were compared by two-way ANOVA ($p < 0.0001$). Error bars represent standard deviation.

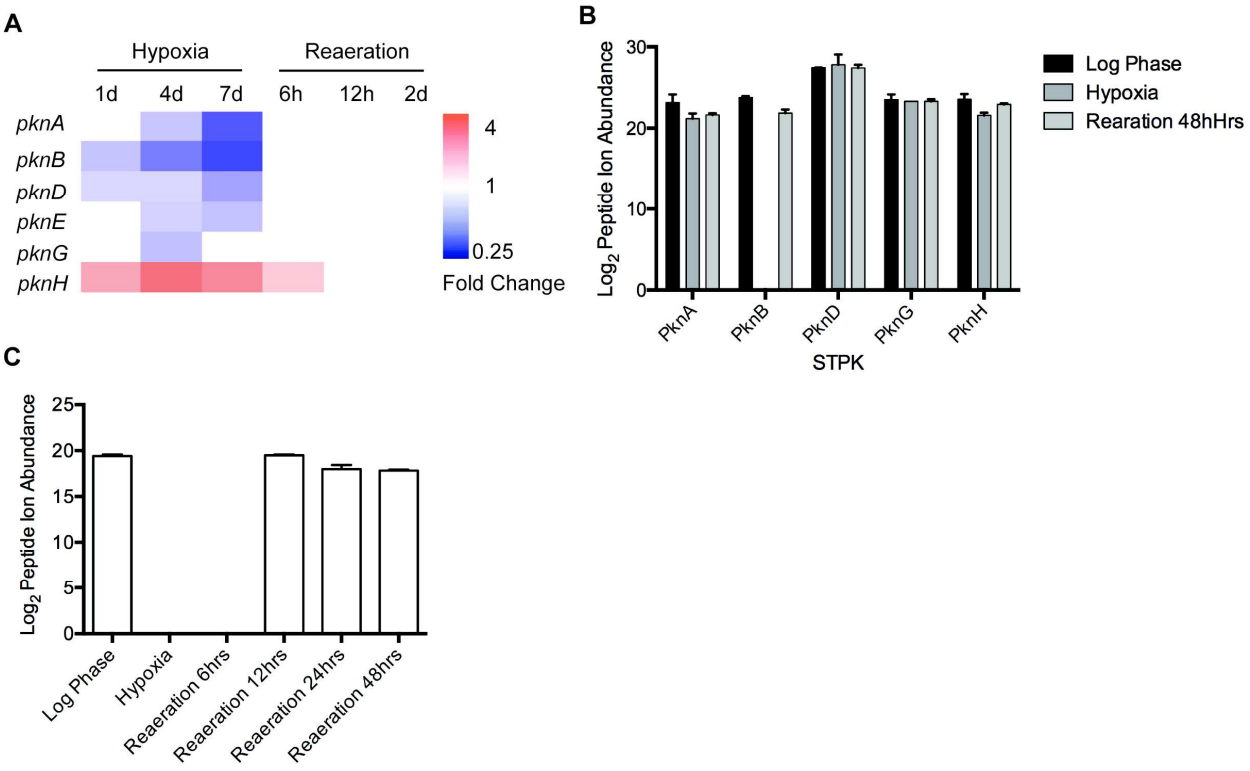


Figure 3-4. PknB transcript and protein levels change in response to oxygen. (A) Microarray analysis of STPK transcript levels during a hypoxia time course. (B) Quantitative MS analysis of STPKs in wild-type *Mtb* in log phase, day 7 of hypoxia, and day 2 of reaeration. (C) Quantitative MS analysis of PknB abundance in wild-type *Mtb*.

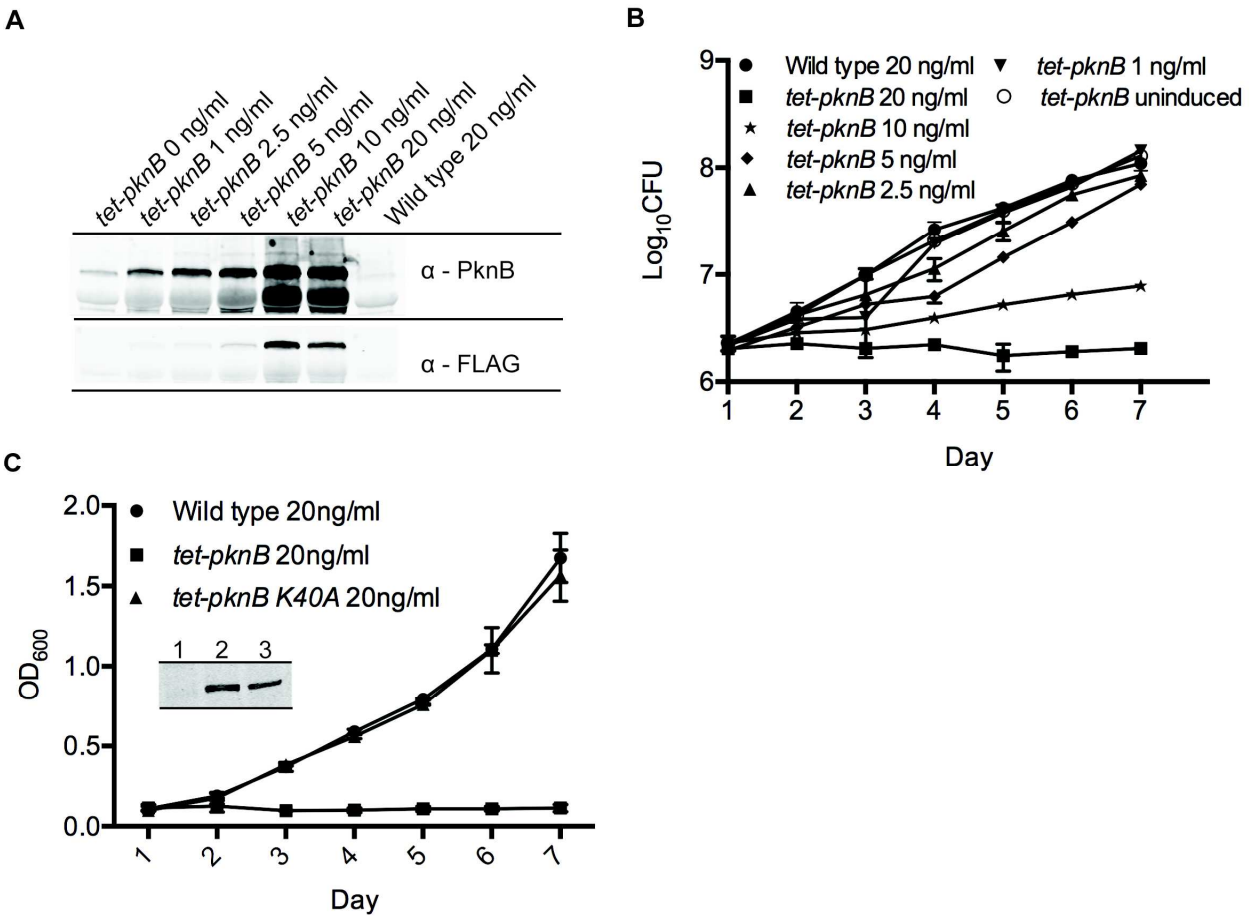


Figure 3-5. *Mtb* growth is sensitive to elevated PknB levels. (A) Overexpression of PknB by the *tet-pknB* mutant upon induction with ATc. PknB levels were determined by Western blot using PknB-specific rabbit IgG and anti-FLAG antibody. (B) Effects of PknB overexpression measured by CFU for 6 d post-ATc induction and by (C) optical density in comparison to the kinase dead mutant *tet-pknB K40A*. (Inset) PknB expression in the *tet-pknB* and *tet-pknB K40A* mutants was determined by anti-FLAG Western Blot. 1, wild type; 2, *tet-pknB*; 3, *tet-pknB K40A,* all treated with 20 ng/ml ATc. CFUs were compared by two-way ANOVA ($p < 0.0001$). Error bars represent standard deviation.

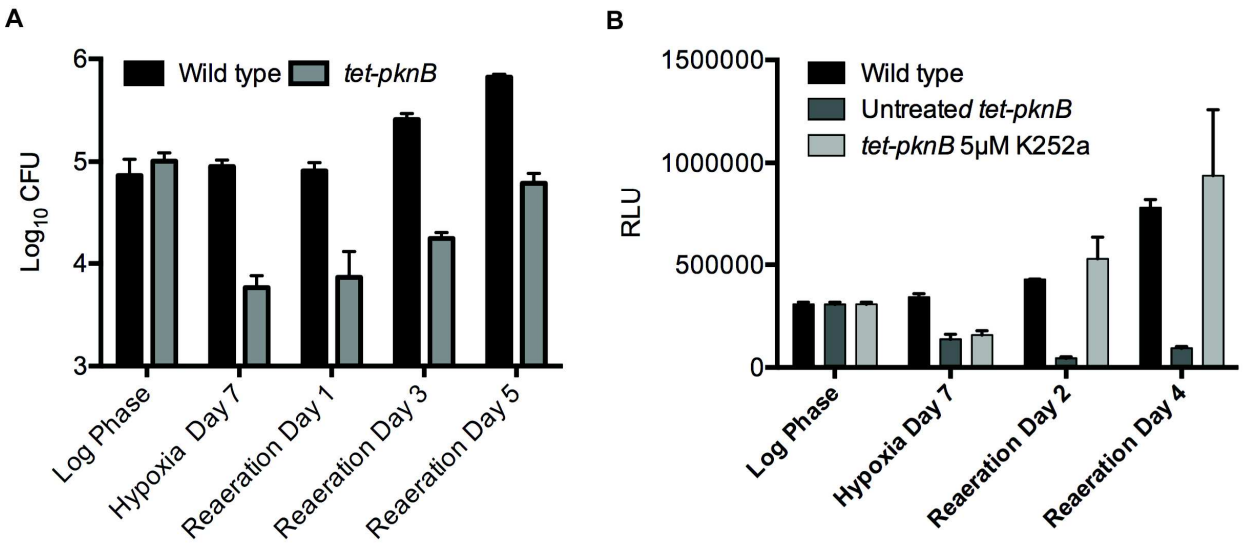
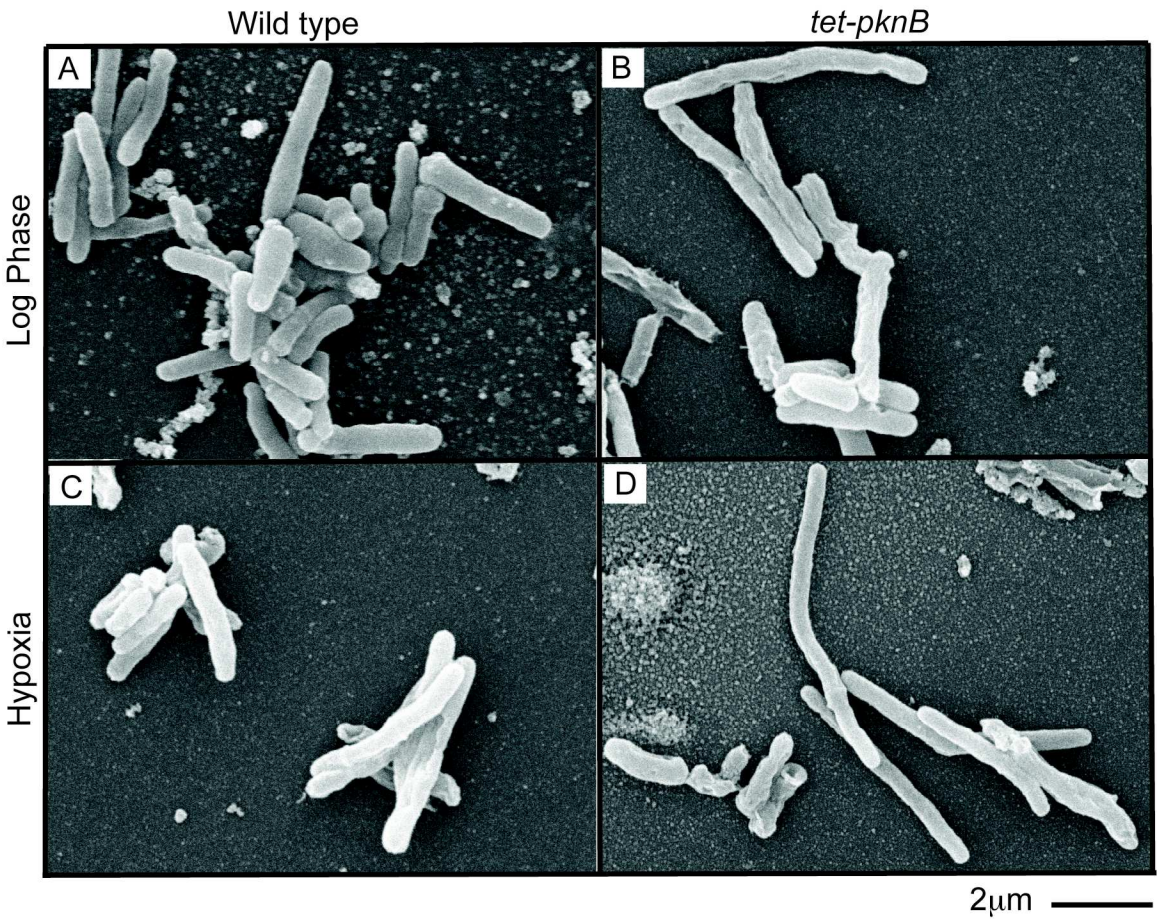


Figure 3-6. Altered PknB levels disrupt oxygen-dependent replication. Viability of (A) uninduced *tet-pknB* and (B) K252a-treated *tet-pknB* was measured by CFU and ATP levels on day 0, hypoxia day 7, and reaeration day 1 and day 3. CFUs and ATP levels were compared by two-way ANOVA ($p < 0.0001$). Error bars represent standard deviation.



E

Log Phase		Hypoxia	
WT	*tet-pknB*	WT	*tet-pknB*
2%	21%	3%	46%

Figure 3-7. PknB overexpression leads to elongated bacilli in hypoxia. Scanning electron microscopy of (A) wild-type and (B) *tet-pknB* on day 0 (log phase), and (C) wild-type and (D) *tet-pknB* after 7 d of hypoxia. (E) The percent of elongated bacilli for each strain in log phase and hypoxia.

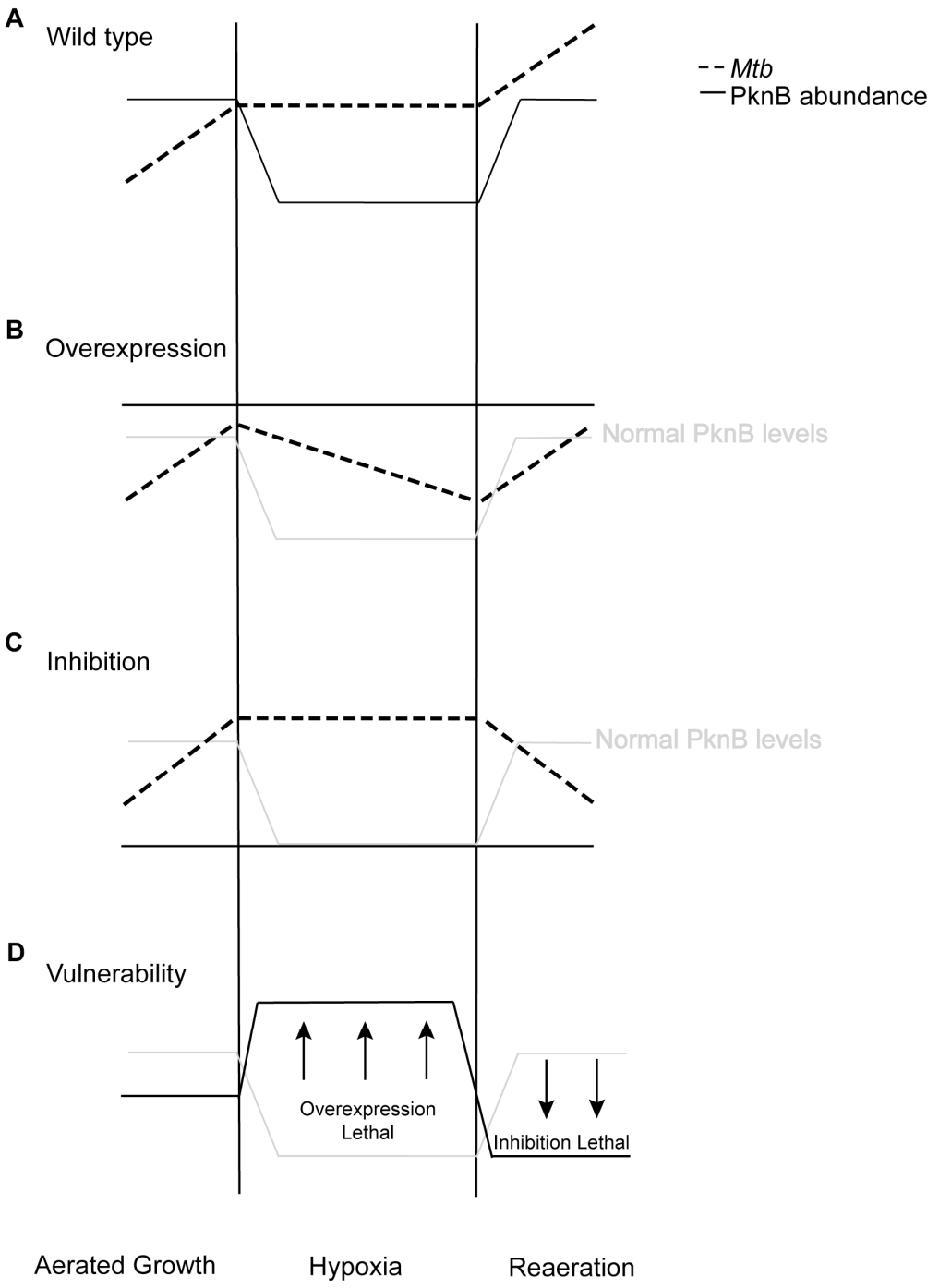


Figure 3-8. Effects of PknB dysregulation in hypoxia and reaeration. (A) *Mtb* growth (dashed line) is regulated by PknB (solid line) in response to oxygen tension. Under low oxygen conditions in hypoxia, PknB activity is decreased to facilitate growth arrest, but returns upon reaeration to support regrowth. PknB dysregulation by (B) overexpression or (C) inhibition interferes with survival. (D) *Mtb* growth is most vulnerable to PknB targeting in hypoxia and reaeration. Elevated PknB levels in hypoxia and reduced levels in reaeration are most detrimental.

Supporting Information

Supporting Figures

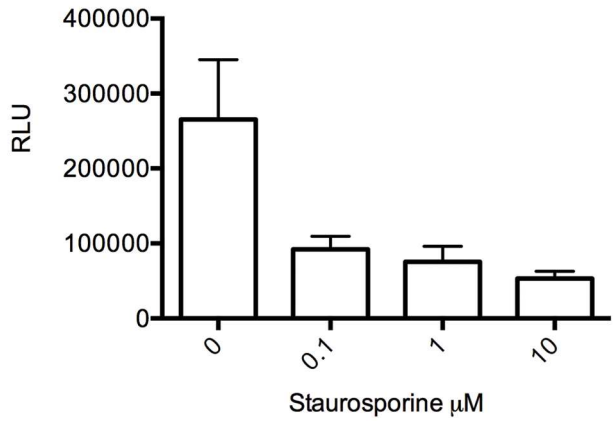


Figure 3-S1. **Inhibition by staurosporine compromises viability in reaeration**. ATP levels were measured using the BTG assay on day 1 of reaeration. ATP levels were compared by one-way ANOVA with Tukey's multiple comparisons test ($p < 0.001$ across all comparisons). Error bars represent standard deviation.

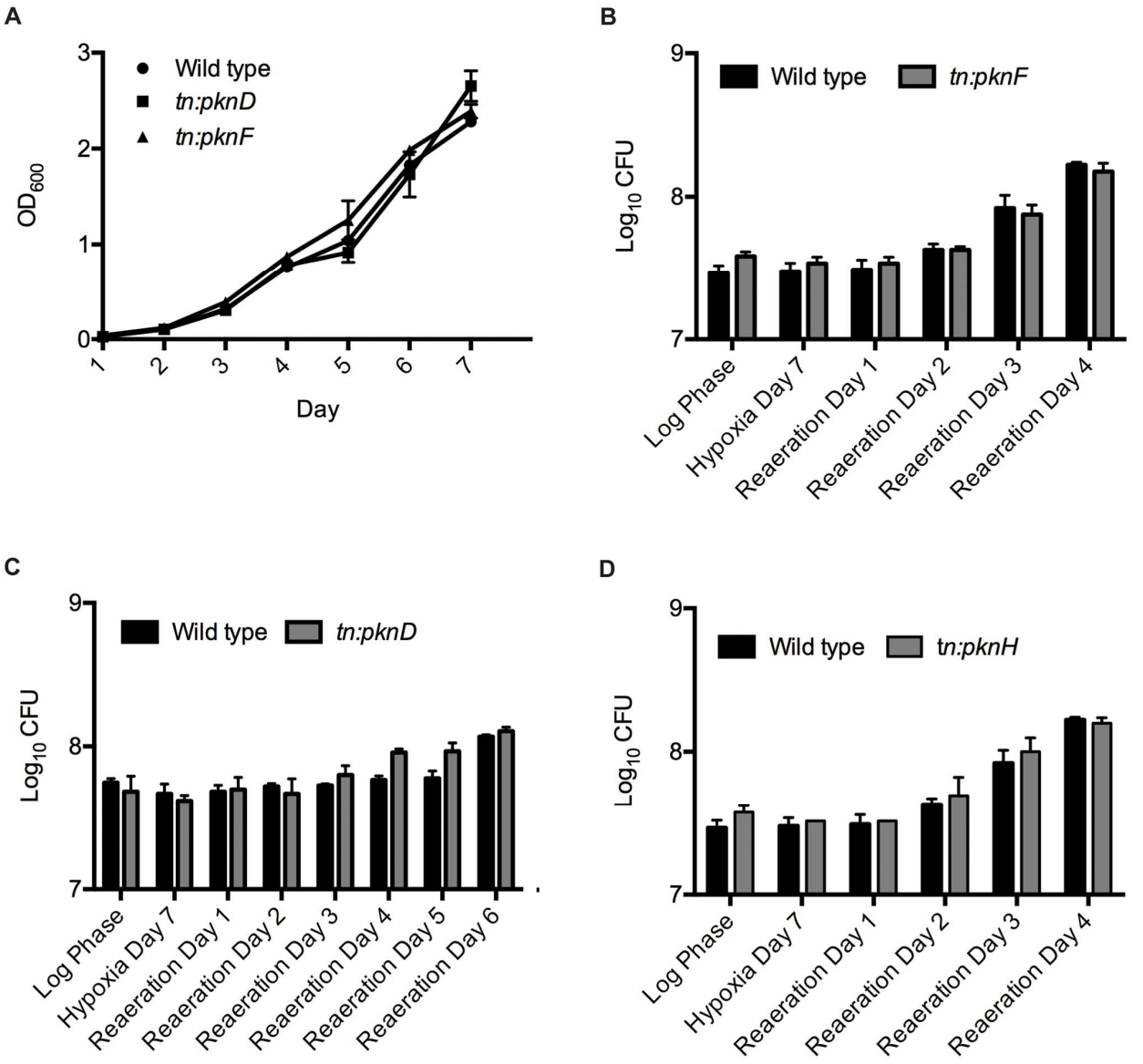


Figure 3-S2. PknD, PknF, and PknH do not have a role in oxygen-dependent replication. (A) Growth kinetics of *tn:pknD* and *tn:pknF* was similar to wild-type. Viability in hypoxia and reaeration, determined by CFU, for (B) *tn:pknD*, (C) *tn:pknF,* and (D) *tn:pknH* was also comparable to wild type. CFUs were compared by two-way ANOVA. Error bars represent standard deviation.

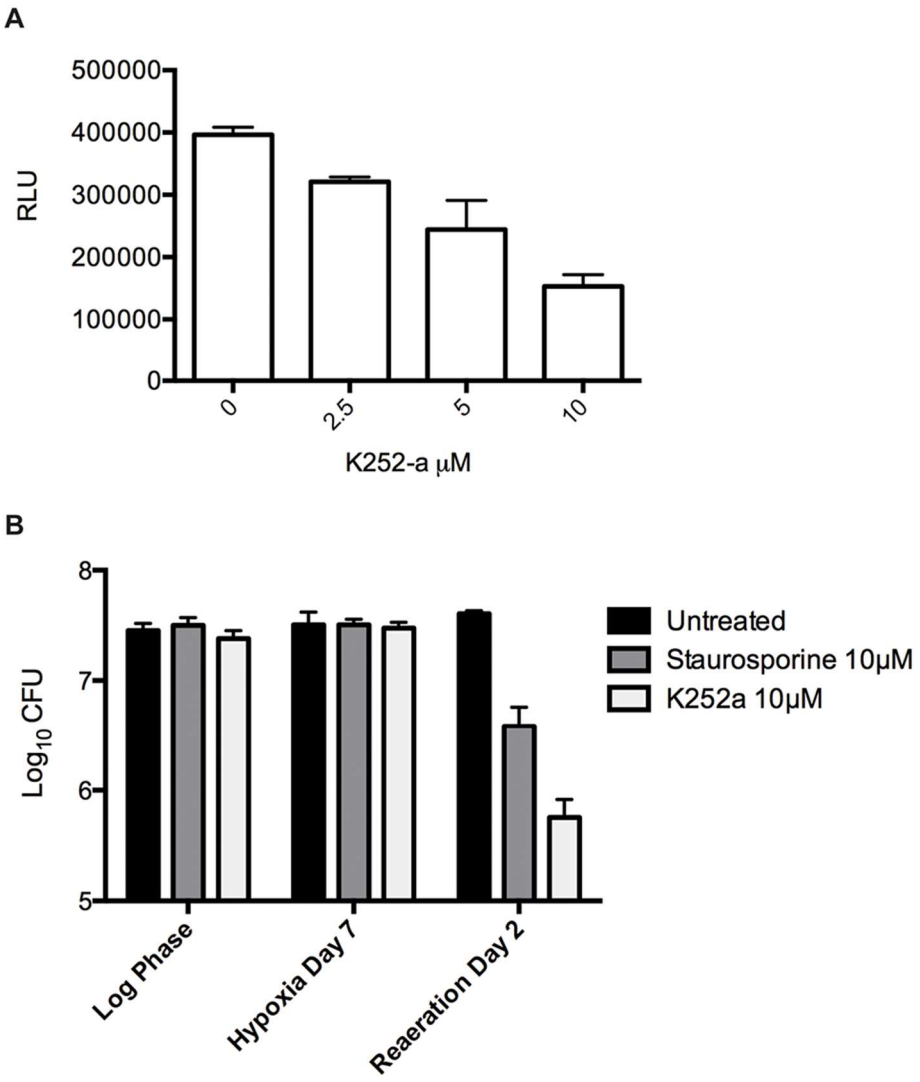


Figure 3-S3. K252a inhibition compromises viability in reaeration. ATP levels were measured using the BTG assay on day 1 of reaeration. ATP levels were compared by one-way ANOVA with Tukey's multiple comparisons test ($p < 0.001$ across all comparisons). Error bars represent standard deviation.

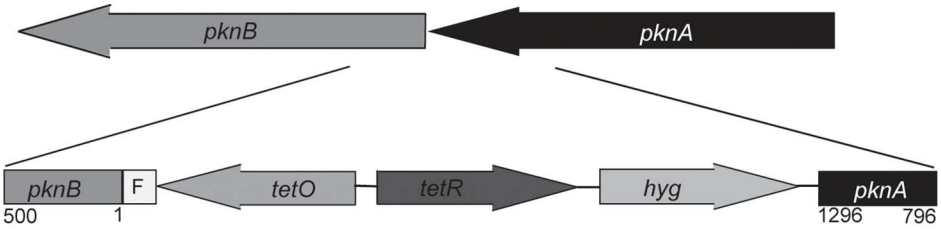


Figure 3-S4. Schematic of the *tet-pknB* overexpression mutant. The recombineering substrate consisted of a hygromycin cassette (*hyg*), the tetracycline promoter elements (*tetR* and *tetO*), and an N-terminal FLAG tag (F), flanked by 500 base pairs of homology to the 3' end of the *pknA* gene (796–1296) and the 5' end of the *pknB* gene (1–500). Electroporation into mycobacteria was performed as described previously [126].

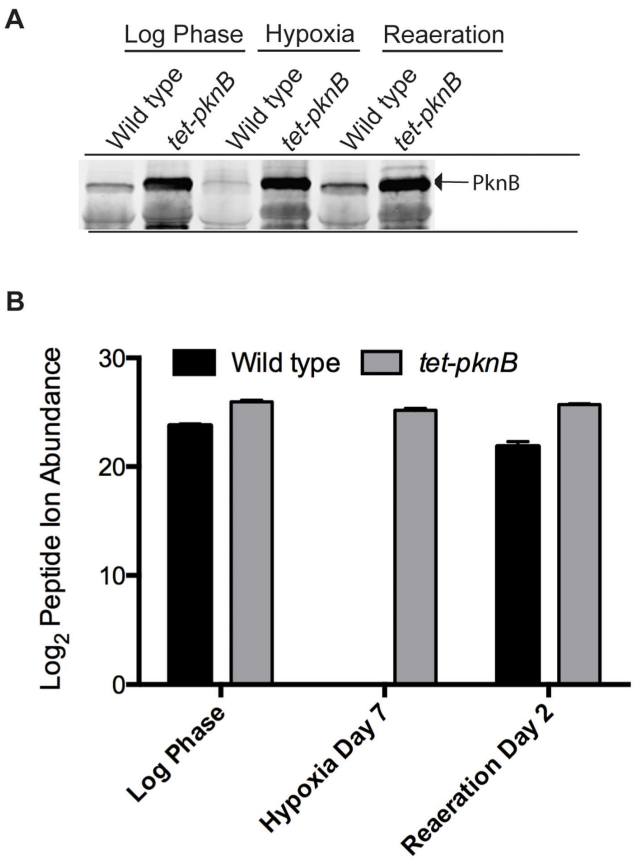


Figure 3-S5. Uninduced *tet-pknB* overexpresses PknB in hypoxia. PknB protein levels measured by (A) Western blot using PknB-specific rabbit IgG and (B) mass spectrometry in wild-type and *tet-pknB* cell lysate during a hypoxia time course.

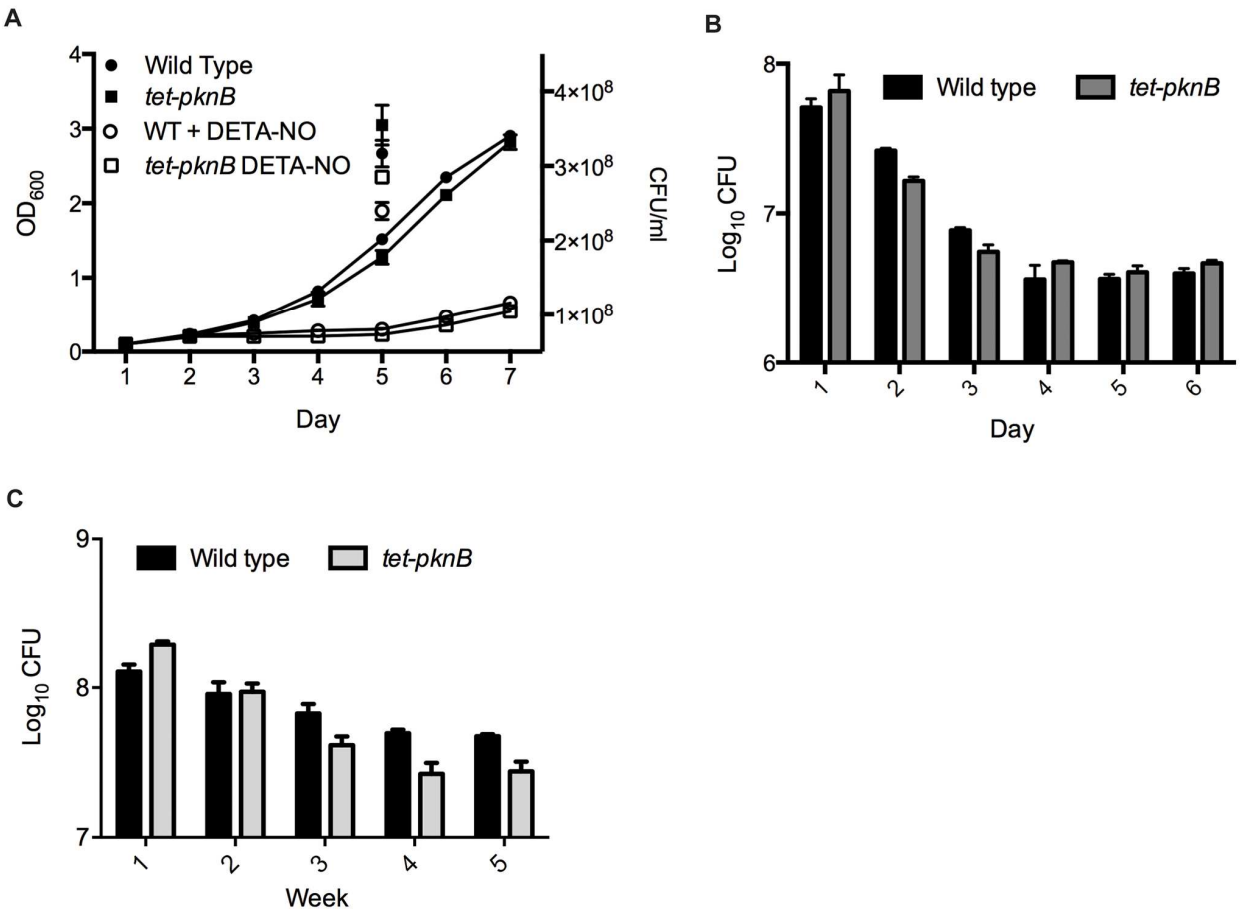


Figure 3-S6. PknB mediates an oxygen-specific replication switch. The *tet-pknB* mutant was exposed to (A) nitric oxide daily for 4 d, (B) pH 4.5 for 5 d, and (C) nutrient starvation for 4 wk. CFUs were compared by two-way ANOVA and not significantly different between wild-type and *tet-pknB*. Error bars represent standard deviation.

Supporting Table

http://figshare.com/articles/_Mycobacterium_tuberculosis_Ser_Thr_Protein_Kinase_B_Mediates_an_Oxygen_Dependent_Replication_Switch/895600

Table 3-S1 Staurosporine binding profile in the *Mtb* proteome. Rv numbers, fold change between staurosporine treated and untreated samples, and peptide ion abundance as log_2 are shown.
Add link

Chapter 4 - *Mycobacterium tuberculosis* Supports Protein Tyrosine Phosphorylation

The following text is from the article: Ulrike Kusebauch, Corrie Ortega, Anja Ollodart, Richard Rogers, David R. Sherman, Robert L. Moritz, and Christoph Grundner. (2014). *Mycobacterium tuberculosis* Supports Protein Tyrosine Phosphorylation. **Proc Natl Acad Sci**. Published ahead of print. Figure numbers have been updated to conform to the formatting of this dissertation; however, the remainder of the text is as submitted.

Abstract

Reversible protein phosphorylation determines growth and adaptive decisions in *Mycobacterium tuberculosis* (*Mtb*). At least 11 two-component systems and 11 Ser/Thr protein kinases (STPKs) mediate phosphorylation on aspartate, histidine, serine, and threonine. In contrast, protein phosphorylation on tyrosine (Tyr) has not been described in *Mtb*. Here, using a combination of phospho-enrichment and highly sensitive mass spectrometry, we show extensive protein Tyr phosphorylation of diverse *Mtb* proteins including STPKs. Several STPKs function as dual specificity kinases that phosphorylate Tyr in *cis* and in *trans*, suggesting that dual specificity kinases have a major role in bacterial phosphosignaling. Mutation of a phosphotyrosine site of the essential STPK PknB reduces its activity *in vitro* and in live *Mtb*, indicating that Tyr phosphorylation has a functional role in bacterial growth. These data identify a previously unrecognized phosphorylation system in a human pathogen that claims ~1.4 million lives every year.

Significance statement

Reversible protein phosphorylation is a major regulatory mechanism by which bacteria sense and respond to changes in their environment. In *Mtb*, however, protein phosphorylation on Tyr residues has not yet been described and is thought to be absent. We show that *Mtb* phosphorylates diverse proteins on tyrosine, suggesting a broad functional role. We identify the

Ser/Thr kinases as the kinases also responsible for phosphorylation on Tyr and show that Tyr phosphorylation regulates Ser/Thr protein kinase activity. Together, our study provides the basis to understand how this new *Mtb* posttranslational modification affects physiology and pathogenesis.

Introduction

Reversible phosphorylation is the main signaling mechanism underlying the dynamic adaptive responses necessary for *Mtb* survival in the host. While phosphorylation on serine (Ser), threonine (Thr), and tyrosine (Tyr) was long considered exclusive to the eukaryotic lineage, it is now recognized as ancient and ubiquitous [39]. In the major human pathogen *Mtb*, phosphorylation on Ser and Thr residues is essential for growth and adaptation, and the essential growth regulator protein kinase PknB is under development as a therapeutic target [111,129]. In contrast, the prevalence and functional consequences of protein phosphorylation on Tyr in prokaryotes are less defined. The best-understood role of bacterial Tyr phosphorylation is the regulation of capsule exopolysaccharide production and transport in several species, directly linking Tyr phosphorylation to pathogenesis [130-133].

In *Mtb*, the phosphosignaling landscape is not well defined, and there is no conclusive molecular evidence for protein Tyr phosphorylation to date. One study detected a small number of bands in *Mtb* lysate by Western blotting with a phospho Tyr (pTyr)-specific antibody (4G10), but these proteins were not identified and phosphorylation on Tyr not confirmed by other methods [59]. Another study showed *in vitro* Tyr phosphorylation of two *Mtb* proteins [58]. Several phosphoproteomic studies in *Mtb* identified hundreds of pSer/pThr sites, but did not detect pTyr sites [48,91]. The prokaryotic Phosphorylation Site Database [134] lists 167 pTyr sites from a number of bacterial species, but none from *Mtb*. Bacterial Tyr phosphorylation is thought to be largely catalyzed by a distinct class of Tyr kinases, the BY kinases, that share no similarity with eukaryotic kinases [135]. *Mtb* is not predicted to encode functional BY kinases,

yet produces two canonical protein Tyr phosphatases of the low molecular weight and classical types [57]. The Tyr phosphatases, however, are thought to be secreted and to act on host substrates [136]. The absence of functional BY kinases and the presumed secretion of the phosphatases into the host added to the current notion that *Mtb* does not support intrinsic protein Tyr phosphorylation.

Phosphorylation on Tyr is the least abundant, most labile, and technically least tractable of Ser/Thr/Tyr phosphorylation events. In eukaryotes, pTyr contributes 0.5-2% of phosphorylation sites [137,138]. In prokaryotes, the abundance of Tyr phosphorylation differs widely between species. Most studies reported between 2.7 and 25.8% of Ser/Thr/Tyr phosphorylation events occurring on Tyr [139], including 8.6% in *E. coli* [140]. However, a recent study in *E. coli* identified >500 pTyr sites, a number far exceeding that of Ser/Thr phosphosites, suggesting much wider prevalence of Tyr phosphorylation than previously believed [51]. The same study suggests a much smaller role of the *E. coli* BY kinases Etk and Wzc than previously thought.

The presence of Tyr phosphorylation in most bacterial phyla, but presumed absence in others led us to reconsider the possibility of Tyr phosphorylation in *Mtb*. Here, we show that *Mtb* does indeed phosphorylate protein on Tyr. We found that several STPKs phosphorylate protein on Tyr in *cis* and in *trans*, and that Tyr phosphorylation regulates STPK activity. Using a gel-free, high mass accuracy mass spectrometry (MS) approach, we identified several Tyr-phosphorylated proteins in *Mtb*, indicating potentially broad regulation of *Mtb* physiology by this posttranslational modification.

Results

Mtb STPKs are Dual Specificity Kinases

Protein phosphorylation is a major regulatory mechanism in *Mtb*. Protein Tyr phosphorylation is widespread among prokaryotes from *Archaea* to *Firmicutes* and

Proteobacteria; but there has yet to be evidence for Tyr phosphorylation in *Mtb*. In contrast to the current notion, we hypothesized that *Mtb* may in fact regulate cellular processes by Tyr phosphorylation. To explore this idea, we first investigated whether *Mtb* encodes bacterial BY kinases, a distinct family of Tyr kinases with no sequence or structural similarity to eukaryotic Tyr kinases. To identify potential BY kinases in *Mtb*, we searched the *Mtb* proteome using the BYKdb prediction program [141]. BYKdb did not predict any functional *Mtb* BY kinases, but identified two proteins with similarity to BY kinases, Rv1708 and Rv3918c. However, both Rv1708 and Rv3918c do not contain the signature BY kinase Walker A' motif. In a chemical biology screen for ATPases [114], we previously identified Rv1708 as a functional ATPase, a prerequisite for kinase function. To verify the prediction that Rv1708 is non-functional although it has ATPase activity, we expressed Rv1708 recombinantly and tested for kinase activity. Rv1708 did not have any detectable kinase activity (Figure S1A). In an *in vitro* study, Rv2232 has previously been shown to phosphorylate protein on Tyr [58]. Rv2232 is predicted to be a haloacid dehydrogenase, a hydrolase family with no similarity to known kinases. Consistent with the structural prediction for Rv2232, we did not detect kinase activity by measuring ^{32}P incorporation (Figure S1A). Also, we did not detect any Tyr-phosphorylated residues of Rv2232 by MS after incubation with ATP and a range of cofactors.

Although our findings suggested that *Mtb* does not produce BY kinases, Tyr phosphorylation might be carried out by non-canonical mechanisms, consistent with a recent study of Tyr phosphorylation in *E. coli* indicating that, despite their biochemical capacity, BY kinases actually play a limited role in Tyr phosphorylation *in vivo* [51]. In yeast, Tyr phosphorylation is exclusively carried out by dual specificity kinases that have both Ser/Thr and Tyr kinase activity. Although rare, dual specificity kinases have also been described in bacteria [55,56], raising the possibility that some *Mtb* STPKs could also have dual specificity. To test this possibility, we measured *in vitro* Tyr phosphorylation of the kinase domain (KD) of the *Mtb* transcriptional regulator kinase PknD. Surprisingly, in addition to several previously described

pSer and pThr sites, we observed MS spectra for phosphorylation on the activation segment residues Tyr178 and Tyr180 (Figure 1A and B), as well as residues outside of the activation segment. Because *E. coli* also phosphorylates protein on Tyr, we sought to rule out that PknD phosphorylation is a result of *E. coli* phosphorylation during recombinant expression by testing for PknD Tyr autophosphorylation activity. PknD showed clear autophosphorylation on Tyr upon incubation of recombinant PknD with ATP and manganese by Western blotting (Figure 1C), confirming that PknD autophosphorylates on Tyr and can act as a dual specificity kinase. Mutation of Tyr178 and Tyr180 to Phe abrogated autophosphorylation, showing that these sites are indeed phosphorylated (Figure 1C).

Activation segment phosphorylation is a central regulator of kinase activity. To gauge the possibility for Tyr phosphorylation on other *Mtb* STPK activation segments, we aligned the activation segments of the eleven *Mtb* STPKs (Figure 1D). All but one activation segment contained Tyr residues, including the Tyr residue towards the C-terminal end that is conserved in some STPK families and present in all *Mtb* STPKs except PknG. Several other Tyr residues are in the vicinity of pThr sites known to control STPK activation. To test for Tyr phosphorylation in the STPK activation segment sequences, we phosphorylated recombinant PknB, PknD, PknE, PknF, and PknH KDs and PknG full length *in vitro* and investigated Tyr phosphorylation by MS/MS. All tested STPKs were phosphorylated on Tyr (Figure 1D, Table S1), suggesting that most if not all *Mtb* STPKs have dual specificity. PknB, D, E, F, and G were phosphorylated on Tyr in the activation segment, suggesting that several STPKs may also be regulated by Tyr phosphorylation. To rule out that Tyr phosphorylation of these STPKs occurs during recombinant expression as a result of *E. coli* Tyr kinases rather than intrinsic Tyr kinase activity, we compared Tyr phosphorylation on kinase dead PknB, F, and H with wild type - all Tyr phosphorylation on the kinase dead mutants would be the result of *E. coli* phosphorylation. We detected no or one (PknH) pTyr on the dead kinases, but numerous pTyr on the WT, indicating

that the pTyr detected in recombinant STPKs are indeed due to STPK Tyr kinase activity (Figure S1B).

The STPKs Phosphorylate Tyr in *trans*

To test whether the STPKs phosphorylate Tyr in trans, we mapped the Tyr phosphosites on structural models predicted by the PHYRE Protein Fold Recognition Server. Three pTyr sites of PknD mapped to the backside of the N-lobe, and one site to the C-lobe. Most of these pTyr residues are oriented in the opposite direction or are >10 Å away from the predicted position of the ATP γ-phosphate, precluding intramolecular autophosphorylation of these sites. PknH was Tyr-phosphorylated at residues 16 and 31 in the N-lobe, which are also distant and oriented away from the active site. We detected similar phosphorylation sites that are not accessible for intramolecular autophosphorylation in PknB and PknE. These data suggest that the STPKs can phosphorylate substrates in *trans*.

***Mtb* Phosphorylates Proteins on Tyr**

Given our finding that the *Mtb* STPKs can phosphorylate on Tyr in *trans*, we next investigated whether *Mtb* supports broader *in vivo* protein phosphorylation on Tyr. By using a targeted, gel-free, mass spectrometry-based proteomics approach in combination with affinity-based phosphopeptide isolation, we probed the *Mtb* phosphoproteome specifically for phosphorylation on Tyr. *Mtb* cultures were grown to late log phase, lysed, proteolytically digested and subsequently enriched for phosphopeptides by tandem immobilized metal affinity chromatography using Fe(III) and TiO_2 procedures to reduce sample complexity. We performed high-mass accuracy nano liquid-chromatography-mass spectrometry analysis with collision-induced dissociation (LC-MS/MS) of the isolated analytes to comprehensively study the phosphoproteome. Data were analyzed with the proteomics software tool suite Trans-Proteomic-Pipeline (TPP) [142]. We detected 30 unambiguous Tyr phosphorylation sites on 17 proteins in *Mtb* lysates (Table S2). A representative fragmentation spectrum of the Tyr-phosphorylated peptide VPGpYAPQGGGYAEPAGR derived from the protein FhaA is shown in

Figure 2A. Nearly the entire y- and b-ion series in the spectrum could be annotated. The two most intense fragments, $y12^{+}$ and $y16^{++}$, result from internal proline fragmentation in the peptide sequence, a well-known collision-induced fragmentation behavior of this amino acid [143]. Less common but not unusual is a doubly charged fragment derived from a doubly charged precursor such as the $y16^{++}$ ion which may occur due to a mobile proton [143]. The probability of phosphorylation on Tyr388 was determined as 1.000 by the PTMProphet algorithm of the TPP [142]. To further verify Tyr phosphorylation at position 388 and others identified in *Mtb* lysate, we chemically synthesized the FhaA phosphopeptide as well as a subset of the phosphopeptides listed in Table S2 and subjected the synthetic peptides to identical LC-MS/MS analysis. The spectrum obtained from the synthetic Tyr-phosphorylated peptide VPGpYAPQGGGYAEPAGR (Figure 2B) is identical to the fragmentation pattern of the endogenous peptide observed in *Mtb* cell lysate and provides conclusive evidence of the pTyr assignment. The MS/MS spectrum of the non-phosphorylated peptide is shown as comparison (Figure 2C). The observed b and y fragment ions for the phosphorylated and non-phosphorylated form of this peptide are annotated in the peptide sequence. The mass shift due to phosphorylation is shown by the different y ion masses upstream, and the different b ion masses downstream of the pTyr residue, as highlighted by arrows (Figure 2D). Spectra of additional synthetic pTyr peptides were also nearly identical to the spectra obtained from the endogenous peptides in *Mtb* lysate, confirming Tyr phosphorylation of these sites (Figure S2). In addition, we fragmented peptides from *Mtb* lysate by higher energy collisional dissociation in an Exactive mass spectrometer. The low molecular weight phosphotyrosine-specific immonium reporter ion of *m/z* 216.0426 [144] was observed as shown in a representative spectrum in Figure S3 and further confirms the presence of phosphotyrosine in these samples.

In five biological replicates, we detected 30% of phosphorylation sites multiple times, indicating non-stochastic, specific phosphorylation events (Table S2). The overlap between different biological replicates was similar to that reported for analyses of eukaryotic phosphoproteomes [145]. The *Mtb* proteins phosphorylated on Tyr span all functional categories except for the PE/PPE family, and essentials are enriched by ~2-fold compared to all genes (Table S2). The most highly phosphorylated protein both by number of different sites and total abundance is FhaA. FhaA is a forkhead-associated (FHA) domain-containing protein that is also a substrate of several Ser/Thr kinases [146]. Of note, FhaA is a 57kD protein and thus likely the major protein detected by anti-pTyr Western blot in a previous study [59]. All FhaA phosphorylation sites mapped to the unstructured intermediate domain of the protein that is Pro, Gly, and Tyr-rich [147]. Based on structure predictions using PHYRE, other Tyr phosphorylation sites are predicted to be in well-structured regions, precluding a general preference for *Mtb* Tyr phosphorylation of disordered regions. Other Tyr-phosphorylated proteins with known or predicted functions include the chaperone GroEL2, the proteasomal subunit PrcB, and the probable aminopeptidase PepB, raising the possibility that protein folding and degradation are regulated by Tyr phosphorylation. Consistent with our *in vitro* data, we also detected phosphorylation of Tyr178 of PknD, showing that STPK activation segment phosphorylation occurs in live *Mtb*.

To determine the ratio of pTyr to pSer and pThr in *Mtb*, we compared the number of Tyr phosphorylation sites to the number of pSer/pThr sites detected in the same samples. In agreement with previous studies of bacterial Ser/Thr/Tyr phosphorylation, we found that 5.8% of *Mtb* proteins are phosphorylated, with a strong bias towards pThr, (ratio of pSer:pThr:pTyr of 34%:62%:4%). With 4%, the percentage of pTyr in *Mtb* is within the range found in other bacterial phosphoproteomes [139,148].

Protein Tyr Phosphorylation in Published *Mtb* Proteomic Data

In light of our identification of Tyr phosphorylation *in vitro* and *in vivo*, it is puzzling that numerous previous studies did not detect this modification. To test if phosphorylated Tyr peptides may also be contained albeit overlooked in previous data, we analyzed four data sets available through the public MS data repositories PeptideAtlas (www.peptideatlas.org) and Peptidome (www.ncbi.nlm.nih.gov/peptidome). We considered two recently published data sets on *Mtb* strain H37Rv described by Kelkar *et al.* [149] (123 LC-MS/MS runs, 1385970 spectra) and Schubert *et al.* [150] (24 LC-MS/MS runs, 236475 spectra), one *M. bovis* BCG set by Schubert *et al.* [150] (24 LC-MS/MS runs, 213183 spectra) and one set on *Mtb* strain A7494 and A12998 (250 LC-MS/MS runs, 1784650 spectra, Peptidome PSE133). Each study used different types of sample fractionation including OffGel electrophoresis, strong cation exchange chromatography, C18 chromatography, MudPIT or in-gel fractionation of total cell lysate and culture filtrate. Due to the different scope of each study and the assumption that *Mtb* does not support phosphorylation on Tyr, to the best of our knowledge, none of these data were extensively analyzed with regard to Tyr phosphorylation. The search of the more than 3.6 million spectra from these four datasets resulted in the identification of 35 Tyr-phosphorylation sites from 34 proteins from a broad range of functional categories (Table S3). Two peptides from FhaA (R.GGYPPETGGYPPQPG**pY**PRPR.H, R.VPG**pY**APQGGGYAEPAGR.D) were identified both in the public dataset and in our experiments, further confirming assignment of this pTyr site. Of note, former peptide was identified in each of the four public data sets as well as in our study. Rv0440 was also identified through different pTyr peptides in our experiments and the public data. All other phosphorylation sites were different between our and published datasets, likely a result of large differences in sample type, sample preparation, fractionation techniques used, MS approaches and instrumentation, and data analysis. These data demonstrate the utility of re-analysis of public datasets for addressing new biological questions, further expand

our view of the Tyr phosphoproteome, and bring the total number of *Mtb* Tyr-phosphorylated sites to 63 in 49 different proteins.

To test if phosphorylation on Tyr is conserved in other mycobacteria, we also measured pTyr in the soil dwelling *M. smegmatis*. Indeed, we identified four proteins (Table S4) including MSMEG_5785, the *Mtb* orthologue of Rv2145c, which was also identified in the public data set as a Tyr-phosphorylated protein. These data show that Tyr phosphorylation is a conserved regulatory mechanism in mycobacteria. Because *M. smegmatis* is a soil-dwelling, non-pathogenic species, pTyr does not seem to be exclusively linked to pathogenesis, but is likely to support basic bacterial functions.

Phosphosite Mutation Affects PknB Activity

Kinases are widely regulated by phosphorylation in their activation segments. To test the effect of Tyr phosphorylation of the activation segment on kinase activity, we mutated the PknB KD pTyr site 182 to Phe and generated a Thr173Ala mutant, which is known to regulate PknB activity [120]. All proteins were purified by metal affinity and size exclusion chromatography and eluted at a similar molecular weight (Figure S4A). We tested for autophosphorylating activity by measuring ^{32}P incorporation after incubation of PknB with [γ-^{32}P] ATP. Unlike the PknB Thr173Ala mutation, the autophosphorylating activity of Tyr182Phe mutation did not differ from that of wild type (Figure S4B), suggesting that activation segment Tyr phosphorylation is not required for this STPK's autophosphorylating activity. However, trans-phosphorylation of FhaA and GarA, two known PknB substrates that are phosphorylated on Ser and Thr residues [146,151], was reduced in the Tyr182Phe mutant (Figure 3A). PknB is essential for *Mtb* growth, and we recently showed that growth is highly sensitive to altered PknB activity [152]. To test whether the reduced activity of the Tyr182Phe mutant *in vitro* affects live *Mtb*, we measured growth of *Mtb* strains overexpressing wild-type PknB, kinase-dead Lys40Ala mutant, and Tyr182Phe mutant (Figure 3B, inset). As previously shown, overexpression of wild-type PknB led to complete growth arrest, whereas the kinase dead mutant did not affect *Mtb* growth.

Overexpression of Tyr182Phe led to a >20-fold reduction of toxicity when compared to overexpression of wild-type PknB on day 7 (Figure 3B). These data show that PknB Tyr182Phe function is impaired in live *Mtb*, suggesting that Tyr phosphorylation of the activation segment regulates some essential functions of PknB relating to growth.

Collectively, we present multiple lines of conclusive evidence for Tyr phosphorylation in *Mtb*, providing a roadmap for defining new Tyr phosphorylation-dependent regulatory mechanisms in this major human pathogen.

Discussion

Protein Tyr phosphorylation in prokaryotes has long been elusive, and in *Mtb*, phosphorylation on Tyr was thought to be absent. Both phosphoproteins and functional Tyr kinases were considered absent to date, and the orphan Tyr phosphatases were plausibly explained as virulence factors targeting host proteins. We now show the first conclusive evidence for protein Tyr phosphorylation in *Mtb*. Along with the recent identification of >500 pTyr sites in *E. coli* [51], these findings suggest more pervasive presence of Tyr phosphorylation in prokaryotes and reveal a new layer of *Mtb* phosphosignaling.

Until recently, most pTyr activity in bacteria was ascribed to BY kinases. However, we found no evidence of functional BY kinases in *Mtb* but instead identified Tyr phosphorylation activity of a number of STPKs, providing a set of candidate kinases that may mediate selective Tyr phosphorylation *in vivo*. Individual bacterial dual specificity kinases have previously been described in *Chlamydophila* and *Bacillus*, but our data suggest that dual specificity kinases might play a prominent role in bacterial phosphosignaling. The BY kinases, on the other hand, may have a more peripheral role in Tyr phosphorylation than previously thought. This idea is consistent with a limited role of the two known *E. coli* BY kinases Etk and Wzc in *E. coli* Tyr phosphorylation: Deletion of both kinases had a small effect on overall Tyr phosphorylation, suggesting the presence of additional non-canonical Tyr kinases in *E. coli* [51].

Our data suggest that Tyr phosphorylation of the activation segment may regulate STPK activity in a way that is reminiscent of other dual specificity kinases. Mutation of the PknB Tyr site 182 led to loss of function *in vitro* and in live *Mtb*, similar to the loss of function previously described for deletion of the Thr173 phosphosite. However, we cannot rule out that the mutagenic removal of a hydroxyl group from Tyr182 has other, phospho-independent effects on activity or substrate binding. Because PknB and PknD activation segment Tyr mutations did not affect autophosphorylation, these Tyr phosphosites might function in other ways, for example by recruiting specific substrates. Our finding of intrinsic Tyr phosphorylation in *Mtb* also puts the Tyr phosphatases PtpA and PtpB in a new light. Although dephosphorylation of host substrates by PtpA has been substantiated, a primary role of PtpA in intrinsic dephosphorylation now appears likely. PtpA deletion does not affect *in vitro* growth in rich medium [153], but causes attenuation of *Mtb* in activated human macrophages [7], pointing to an intracellular role of PtpA in processes that affect host-pathogen interactions. With Tyr phosphorylation regulating the cell wall in other bacteria and with cell wall enzymes phosphorylated on Tyr in *Mtb*, the effect of PtpA on the survival in the host may also relate to control of cell wall biosynthesis.

We identified Tyr phosphorylation sites on proteins with a wide range of functions. Consistent with a role of Tyr phosphorylation in capsule biosynthesis in other bacteria, two essential *Mtb* enzymes generating cell wall-associated polysaccharides are also phosphorylated on Tyr: GlfT2, which catalyzes the polymerization of arabinogalactan with peptidoglycan to form an essential component of the mycobacterial cell wall [154], and the 1,4-alpha-glucan branching enzyme GlgB (Rv1326c [155]) that is involved in the biosynthesis of glycogen and capsular alpha-D-glucan. These data suggest that the regulation of cell wall-associated carbohydrates by Tyr phosphorylation might be conserved in *Mtb*. The most highly Tyr-phosphorylated protein, FhaA, is an FHA-domain containing protein that is also a substrate of several Ser/Thr kinases. A recent study showed a role for FhaA in assembling a signaling complex with PknB and Rv3910 that affects cell wall synthesis [110]. Interestingly, FhaA contains a long intermediate domain in

addition to the C-terminal FHA domain and an N-terminal domain of unknown function. The intermediate region has a highly repetitive, Tyr-rich sequence that is unstructured. All Tyr phosphorylation sites mapped to this intermediate domain. The FhaA intermediate domain could link FHA-mediated functions to other pathways by serving as a switchboard through combinatorial Tyr phosphorylation. In addition, FhaA and several other Tyr-phosphorylated proteins are also phosphorylated on Ser and Thr, showing that both modifications can decorate the same protein simultaneously.

The detection of Tyr phosphorylation in *Mtb* appears to be more challenging than in other organisms. We found that pTyr-specific antibodies cannot readily detect Tyr phosphorylated proteins in cell lysate, and their reactivity even with recombinant Tyr-phosphorylated protein is limited. However, our finding of Tyr phosphorylation sites in existing MS datasets using vastly different fractionation methods shows that the current absence of evidence for Tyr phosphorylation was not primarily due to technical difficulties.

The next challenge for understanding the role of Tyr phosphorylation in *Mtb* is the identification of specific functions of Tyr phosphorylation events. Over 20 years after the discovery of Tyr phosphorylation in bacteria, this study redefines phosphosignaling in *Mtb* and should encourage research into the function of this new arm of signal transduction.

Experimental Procedures

Cloning, expression, and purification of recombinant *Mtb* proteins

The genes for STPK kinase domains, PknG full length, Rv1708, and Rv2232 were amplified from genomic *Mtb H37Rv* DNA and cloned into the pET28b expression vector in-frame with an N-terminal six-histidine tag and a SUMO tag for protein stability (FhaA). The kinase domains contain the following residues: PknB aa1-307, PknD aa1-378, PknE aa1-289, PknF aa1-301, and PknH aa1-399. Site-directed mutagenesis was carried out using the QuikChange protocol (Stratagene). Proteins were expressed and purified as described previously [114].

Western blotting

Kinase reactions were carried with 200-500 ng kinase, 150mM NaCl, 20 mM Tris pH 7.5, 5% glycerol, 1 mM $MnCl_2$ (1mM each of $MnCl_2$ and $MgCl_2$ for Rv1708 and Rv2232), 500 µM Tris (2-carboxyethyl) phosphine hydrochloride, and 1mM ATP. The reactions were incubated for the indicated time at 37°C and stopped by adding SDS-PAGE sample buffer to a final 2-Mercaptoethanol concentration of 1%. The samples were separated by SDS-PAGE, transferred to nitrocellulose membrane, and probed with P-Tyr-100 antibody (Cell Signaling Technology). The blots were probed with secondary antibody Goat Anti-Mouse IRDye® 800cw (Licor) and scanned using the Odyssey digital imagining system.

^{32}P incorporation kinase assay

Kinase reactions were carried out as above, using the indicated amounts of kinase, 5 µg GarA or FhaA substrate, 50 µM ATP, and 5 µCi [γ-^{32}P] ATP per reaction. The reactions were incubated for the indicated time at room temperature and stopped by adding SDS-PAGE sample buffer. The samples were separated by SDS-PAGE, the gels dried, and the ^{32}P incorporation visualized by autoradiography.

***Mtb* strains, culture, and lysate preparation**

Mtb strain H37Rv (ATCC) was grown in liquid 7H9 medium with 10% OADC in rolling cultures. The overexpressing strains *tet-pknB*, *tet-pknB* Lsy40Ala and *tet-pknB* Tyr182Phe were generated as previously described [152]. For CFU assays, cultures at OD_{600} 0.05 were induced with 20ng/ml ATc, and plated daily for 7 days. For LC-MS/MS analysis, cultures were harvested by centrifugation for 5 minutes at 4000 x g and washed once in phosphate buffered saline. Cells were pelleted and resuspended in 50 mM NH_4HCO_3/50% 2,2,2-Trifluoroethanol. Cells were

lysed by bead beating and inactivated by heat killing. *M. smegmatis* strain mc^2 155 was grown and lysate prepared as described above for *Mtb*.

Tryptic digestion and phosphopeptide enrichment

Proteins were reduced with 5mM Dithiothreitol (DDT, 30 min, 55°C), alkylated with 14 mM iodoacetamide (IAM, 30 min, room temperature, darkness), followed by quenching of unreacted IAM with 5 mM DTT. The sample was diluted with 125 mM NH_4HCO_3 and digested with trypsin at 1:50 enzyme:substrate ratio at 37℃ overnight. The digest was dried under centrifugal evaporation (Savant), resolubilized in 500 µl 1% TFA and peptides desalted with tC18 SepPak cartridges (Waters, Milford, MA). To enrich for phosphopeptides, we performed immobilized metal affinity chromatography (IMAC) using PHOS-Select Iron Affinity Gel (Sigma-Aldrich, Saint-Louis) following a similar protocol [156] and then the Titansphere™ Phos-TiO kit (GL Sciences Inc, Tokyo). Prior to MS analysis, peptides were desalted with a tC18 SepPak cartridge. Alternatively, Tyr-phosphorylated proteins were purified by immunoprecipitation with a 1:1 mixture of the anti-pTyr antibodies PY99 agarose and 4G10 Sepharose before tryptic digestion. After three washes in lysis buffer, protein was eluted with 0.2% trifluoroacetic acid, and samples further processed as described above. Recombinant proteins were reduced with 1 mM DDT, alkylated with 10 mM IAM and quenched with 5 mM DDT. Samples were diluted 1:1 with 125 mM NH_4HCO_3 and digested with trypsin (1:50) overnight.

LC-MS/MS analysis

Peptides were analyzed either on an LTQ-Velos Orbitrap or an LTQ-Velos Pro Orbitrap Elite (Thermo Fisher Scientific, San Jose, CA) mass spectrometer equipped with a nano liquid chromatography system. Peptide separation was performed on C18 (Dr. Maisch ReproSil-Pur C18-AQ, 120 Å, 3 µm; Ammerbuch-Entringen, Germany) in-house packed capillary columns using 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B) with a gradient from

3% to 25% B in 90 min at a flow rate of 0.3 µl/min. Survey full scan MS spectra were acquired in the mass range *m*/*z* 400–2000 in the Orbitrap analyzer at a resolution of 60,000. The ten most intense ions determined in the survey scan were fragmented by collision induced dissociation (CID) in the LTQ. Dynamic exclusion was enabled. In addition, a portion of phospho-enriched *Mtb* lysate was analyzed on a Q-Exactive (Thermo-Fisher Scientific, San Jose, CA) by higher energy collisional dissociation and nano LC conditions as described above.

Data analysis

Instrument-native data files were converted to mzML or mzXML files using the ProteoWizard msconvert program [157,158]. MS/MS spectra were associated with peptide sequences using SEQUEST (version UW2012.01.2) and a database comprising 3996 protein entries plus common contaminants and a sequence-shuffled decoy counterpart. Peptides were allowed to be semitryptic with up to two internal cleavage sites. The search parameters included a fixed modification of +57.021464 to account for carbamidomethylated cysteines and differential modifications of +15.9949 for oxidized methionines and +79.966331 for phosphorylated serines, threonines and tyrosines. The search results were processed with the TPP (v4.6 rev1) including PeptideProphet [159] and iProphet [160]. Public MS data were analyzed using the TPP with open source SEQUEST called Comet [161] and X!Tandem [162].

Extended Experimental Procedures are provided in the Supporting Information.

Acknowledgements

We thank Zhi Sun and David Shteynberg for expert assistance and Seemay Chou for helpful discussions. This work has been funded in part with federal funds from the National Science Foundation MRI (grant No. 0923536), by the American Recovery and Reinvestment Act (ARRA) funds through grant number R01 HG005805 from the National Human Genome Research Institute, the National Institute of General Medical Sciences under grant No. S10 RR027584 and

2P50 GM076547/Center for Systems Biology, and the Bill and Melinda Gates Foundation Global Health Grant Number 0PP1039684. CG was supported by the Paul G. Allen Family Foundation Grant #8999 and by grant number 10802993 from NIAID. CO is the recipient of an American Society of Microbiology Robert D. Watkins Graduate Research Fellowship and a Bank of America Endowed Minority Fellowship.

Figures

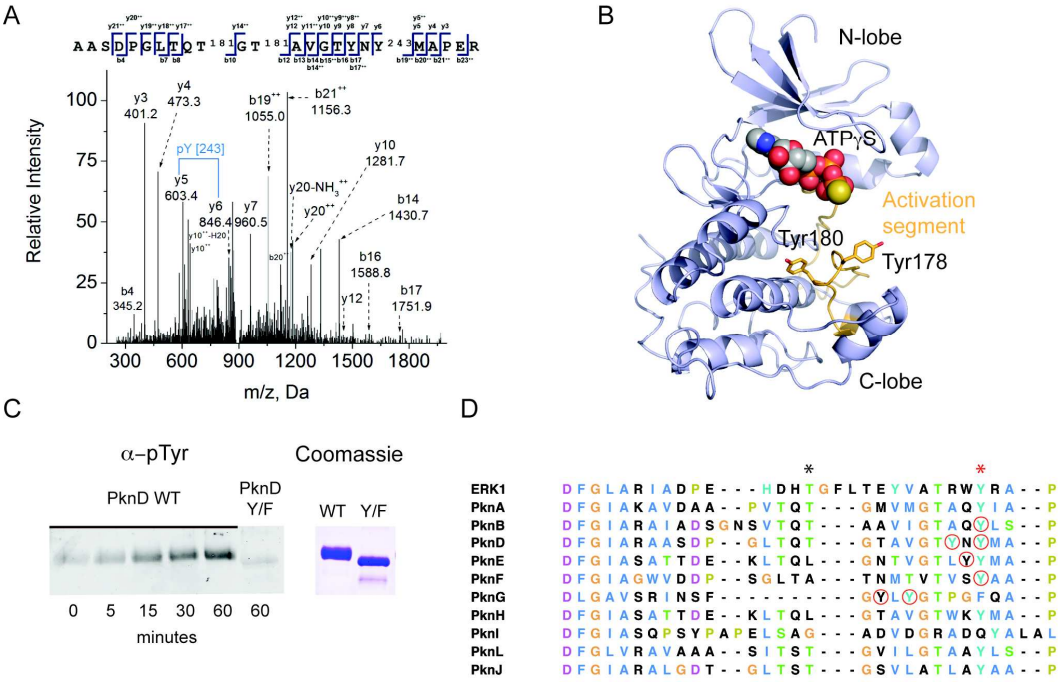


Figure 4-1. **STPKs are dual specificity kinases**. (A) MS/MS spectrum showing phosphorylation of recombinant PknD KD on Tyr180. (B) A PknD structure prediction (PHYRE) shows the predicted position of pTyr178 and 180 on the activation segment (orange) relative to the ATP binding site, N-lobe, and C-lobe. (C) Autophosphorylation of recombinant PknD KD on Tyr. Western blotting with anti-pTyr antibody shows an increase in PknD Tyr phosphorylation after incubation with ATP and Mn^{2+}, and loss of Tyr phosphorylation in the Tyr(178,180)Phe double mutant, showing intrinsic Tyr kinase activity. (D) Sequence alignment of *Mtb* STPK activation segments. Red circles indicate pTyr sites identified in this study. Position of phospho-amino acids that regulate activity in some *Mtb* STPKs is marked by black asterisks, and conserved Tyr by red asterisk.

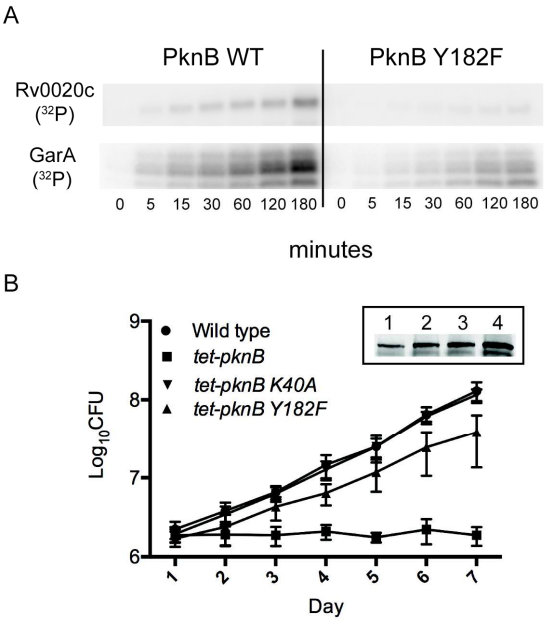


Figure 4-2. **Phosphosite deletion affects STPK activity. (**A) Mutation of Tyr182 to Ala or Phe in PknB reduces PknB activity *in vitro*. Autoradiogram shows loss of phosphorylation of the PknB substrates Rv0020c and GarA upon Tyr mutation. (B) Overexpression of PknB is toxic to *Mtb*, but overexpression of the Tyr182Phe mutant leads to >20-fold decreased toxicity compared to overexpression of wild type PknB. Inset: Western blot of PknB expression in wild type *Mtb* (1), *tet-pknB* overexpressor (2), K40A mutant (3), and Y182F mutant (4). Equal protein loading was controlled by Bradford assay.

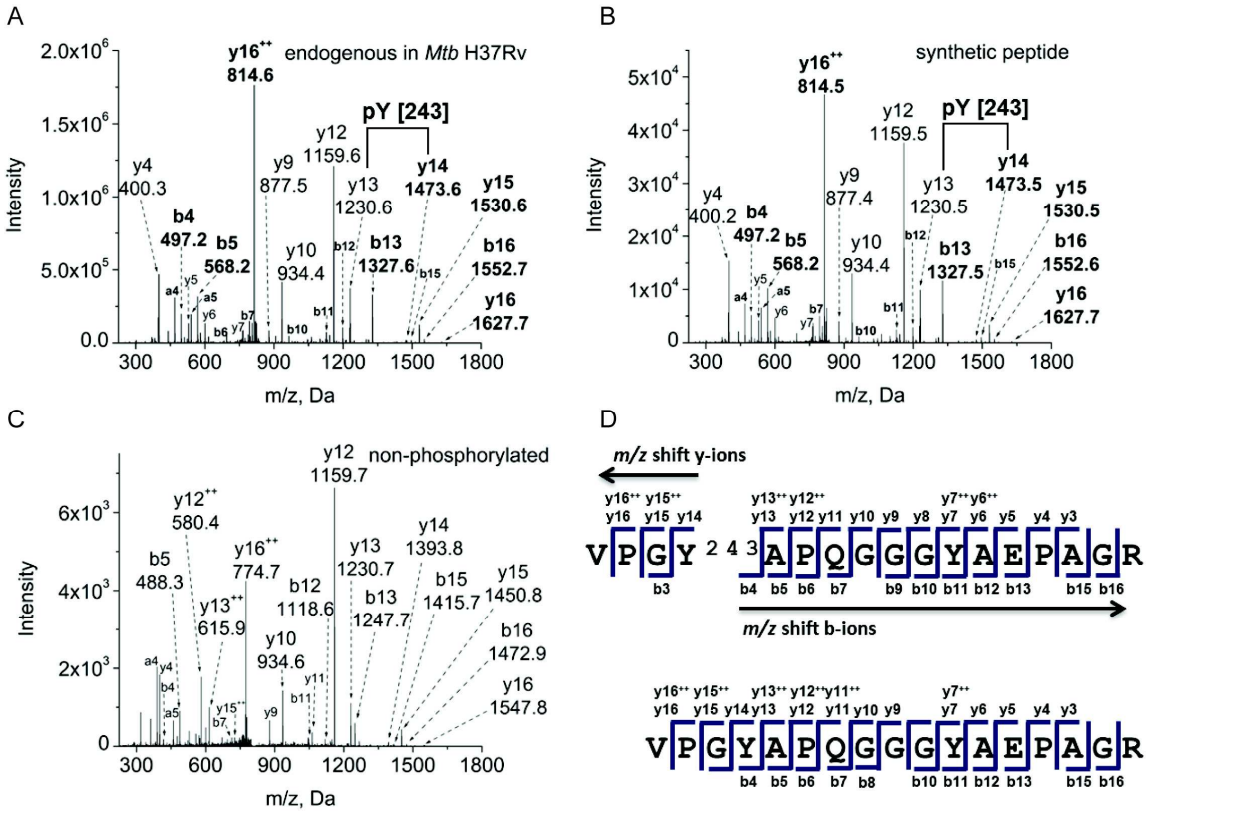


Figure 4-3. **MS/MS spectra of Tyr phosphorylated peptide VPGpYAPQGGGYAEPAGR from Rv0020c.** Precursor ion masses were measured in the Orbitrap Elite mass analyzer and MS/MS spectra were acquired in the LTQ mass spectrometer. (A) Endogenous peptide identified in phosphopeptide-enriched fraction of total *Mtb* cell lysate (precursor ion mass accuracy -0.69 ppm) (B) Synthetic peptide confirming the fragmentation pattern of the endogenous peptide (mass accuracy 1.04 ppm). (C) Non-phosphorylated peptide VPGYAPQGGGYAEPAGR as comparison (mass accuracy 0.61 ppm). (D) Sequence annotation of identified fragment ions of the phosphorylated and non-phosphorylated peptides. The mass shift due to phosphorylation is shown by the different y ion masses observed upstream, and the different b ion masses observed downstream of the pTyr residue, as highlighted by arrows in the phosphorylated peptide sequence and in bold in the spectra.

Supporting Information

Supporting Experimental Procedures

Protein Digestion of Mycobacterium Tuberculosis and Smegmatis

Protein disulfide bonds were reduced with 0.5M Dithiothreitol (DDT) at a final concentration of 5 mM for 30 min at 55°C. After cooling the protein mixture to room temperature, 1 M Iodoacetamide was added to a final concentration of 14 mM to alkylate the sample for 30 min at room temperature in the dark, followed by quenching of unreacted Iodoacetamide with 5 mM DTT for 15 min at room temperature in the dark. The sample was diluted 1:10 with 125 mM NH_4HCO_3, then trypsin (Promega, Madison, WI) was added at 1:50 enzyme:substrate ratio and incubated at 37°C overnight. The digest was dried under centrifugal vacuum evaporation (Savant, Thermo-Fisher, CA), resolubilized in 500 µl 1% trifluoroacetic acid (TFA), acidified to pH<2 and centrifuged at 1,000 g for 5 min to clear the sample. Peptides were desalted with tC18 SepPak cartridges (Waters, Milford, MA). A 100 mg cartridge was washed with 2 ml acetonitrile (ACN), 2 ml 50% ACN / 50% acetic acid (HAcO) and equilibrated with 4 ml 0.1% TFA. The sample was loaded, followed by a washing step with 2 ml 0.1% TFA and 250 µl 0.5% HAcO. Peptides were eluted with 1 ml 50% ACN / 0.5% HAcO and dried under centrifugal vacuum evaporation.

Phosphopeptide Enrichment

To enrich for phosphopeptides we performed immobilized metal affinity chromatography (IMAC) using PHOS-Select Iron Affinity Gel (Sigma-Aldrich, Saint-Louis) following a similar protocol as described by Villen et al. [156] and the Titansphere^{TM} Phos-TiO kit (GL Sciences Inc, Tokyo). 100 µl of PHOS-Select gel were prepared by washing three times with 1 ml 40% ACN / 25 mM formic acid in H_2O (IMAC binding buffer). Dried peptides were reconstituted in 1 ml IMAC binding buffer, mixed with the prepared IMAC beads and incubated with end-over-end rotation

for 45 min. The supernatant was removed and saved for TiO_2 enrichment. IMAC beads were washed with 500 µl IMAC binding buffer, transferred in buffer to a spin column and washed again. Finally phosphopeptides were eluted by incubating for 5 min with 100 µl of 50 mM K_2HPO_4 (IMAC elution buffer, adjusted to pH10 with NH_4OH). Elution was repeated twice, peptides dried under centrifugal vacuum evaporation and desalted as described above. The saved supernatant was dried in a speedvac and subjected to TiO_2 enrichment as described in the Titansphere^{TM}Phos-TiO kit instruction manual. Subsequently peptides were dried by centrifugal vacuum evaporation and cleaned with a tC18 SepPak cartridge.

Synthetic Peptides

Peptides were synthesized using Fmoc chemistry (Thermo Fisher Scientific Huntsville, AL / Ulm, Germany) as free amine at the N-terminus and carboxylic acid at the C-terminus. All cysteine residues were carboxyamidomethylated and phosphorylated tyrosine residues were introduced at the sequence positions as building blocks in the synthesis as they were observed in *Mtb* cell lysates. Peptides were recovered from their synthesis support as crude product, desalted, lyophilized and solubilized in 50% acetonitrile, 0.1% TFA. Peptides were diluted and subjected to LC-MS/MS analysis in 5% acetonitrile/ 0.1% formic acid.

LC-MS Analysis

Peptides were analyzed either on a LTQ-Velos Orbitrap equipped with a nano electrospray ionization source (Thermo Fisher Scientific, San Jose, CA) and a NanoLC-2Dplus HPLC system (Eksigent, Dublin, CA) or on a LTQ-Velos Pro Orbitrap Elite equipped with a nano electrospray ionization source (Thermo Fisher Scientific, San Jose, CA) and a 1100 Series HPLC (Agilent Technologies Inc., Santa Clara, CA) with an electronically controlled flow splitter for nano flow rates. In both cases peptides were loaded to a trap column packed with Dr. Maisch ReproSil-Pur C18-AQ (10x0.075 mm I.D., 120 Å, 3 µm; Ammerbuch-Entringen, Germany) for 5 min using

0.1% formic acid in water at a flow rate of 3 µl/min. Chromatographic separations were performed on C18 in-house packed capillary columns (150x 0.075 mm I.D., Dr. Maisch ReproSil-Pur C18-AQ, 120 Å, 3 µm; Ammerbuch-Entringen, Germany) using 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B) with a gradient from 3% to 25% B in 90 min and 25% to 35% B from 90 - 95 min at a flow rate of 0.3 µl/min. Synthetic peptides were analyzed with a shorter gradient from 3% to 30% B in 30 min and 30% to 60% from 30 – 40 min. Data were acquired with the Xcalibur instrument control software (Thermo Fisher Scientific, San Jose, CA). Survey full scan MS spectra were acquired in the mass range *m*/*z* 400–2000 in the Orbitrap analyzer at a resolution of 60,000. The ten most intense ions determined in the survey scan were fragmented by collision induced dissociation (CID) in the LTQ. Charge state $z = 1$ and unassigned charges were rejected while charge states $z = 2$, 3, and 4 were subjected to MS/MS. MS/MS spectra were acquired upon a minimal signal of 1000 counts, with an isolation width of two and a normalized CE of 35. Dynamic exclusion was enabled to exclude precursors for 60s after a single observation. Dynamic exclusion list size was set to 500 and exclusion mass width to ±10ppm. In addition, some phospho-enriched lysate peptides were analyzed on a Q Exactive equipped with an easy-nLC 1000 HPLC system (Thermo Fisher Scientific, San Jose, CA). Peptide separation was performed as described above. Survey full scan MS spectra were acquired in the mass range *m*/*z* 400–1800 at a resolution of 35,000, with AGC target set to $1e^6$ and maximum IT 150 ms. Peptides were fragmented above a threshold of 1000 by higher energy collisional dissociation (HCD) with a resolution of 17,500, AGC target $1e^5$, maximum IT 100 ms, loop count 30 and isolation width 2.2 Da. Charge state $z = 1$, unassigned charges and $z \geq 6$ were rejected, dynamic exclusion was set to 180 s. Raw mass spectrometry data was deposited in the PeptideAtlas and can be accessed at http://www.peptideatlas.org/PASS/.

Mass Spectrometry Data Analysis

Instrument-native data files were converted to mzML or mzXML using ProteoWizard msconvert [157,158]. MS/MS spectra were associated with peptide sequences using SEQUEST (version UW2012.01.2) and databases provided by the TB Database Project (http://genome.tbdb.org/annotation/genome/tbdb/MultiDownloads.html, a resource through the Stanford School of Medicine, the Broad Institute and the Bill & Melinda Gates foundation). The database comprised 3996 protein entries for *Mycobacterium tuberculosis* and 6716 entries for *Mycobacterium smegmatis*. Common contaminants and a sequence-shuffled decoy counterpart were added to each database. The searches were performed with a mass window of 600 ppm to allow isotopic mass error peptides to be correctly identified within Peptide Prophet [159]. Peptides were allowed to be semitryptic with up to two internal cleavage sides. The search parameters included a fixed modification of +57.021464 to account for carbamidomethylated cysteines and differential modifications of +15.9949 for oxidized methionines and +79.966331 for phosphorylated serines, threonines and tyrosines. The search results were processed with the Trans-Proteomic Pipeline (TPP, v4.6 rev1 and 4.6.2) [142,163] including PeptideProphet, iProphet and PTMProphet [160,164]. Peptide spectrum matches (PSM) generated by the search engine were analyzed with PeptideProphet to assign each PSM a probability of being correct. The accurate mass binning model and the non-parametric model were used in the PeptideProphet analysis. Decoy hits were reported with a computed probability based on the model learned allowing for additional assessment of the performance of the models. The PeptideProphet results were further processed with iProphet to provide a more accurate and conservative representation of the identified peptides resulting in an iProphet probability [160]. The error rate for PSMs was estimated by iProphet for each experiment and only peptides with a minimum iProphet probability corresponding to an error of <0.9% were considered. For the majority of peptides this translates into an iProphet probability >0.9, iProphet probabilities for each peptide are listed in each table. In addition, we applied PTMProphet (available within the TPP) [142] to assign a probability score of the correct phosphorylation site localization.

Synthetic peptides and recombinant proteins were analyzed in the same way. In both cases, the synthetic peptides as well as the set of recombinant proteins, only peptides with a minimum probability of 0.9 were considered, corresponding to an error <0.4% as estimated by iProphet. The sequence coverage of the recombinant proteins was determined by ProteinProphet [142].

For *Mycobacterium smegmatis* spectra were searched with SEQUEST as described above and in addition with X!Tandem (version 2011.12.01.1) [165], OMSSA (version 2.1.8) and Comet (version 2012.01 rev. 3) [161] using similar search criteria. X!Tandem utilized the k-score plugin and the parent monoisotopic mass error was set to 20ppm [163]. For Comet, a peptide mass tolerance of 20ppm was used in combination with the isotope error enabled. The PeptideProphet results from each of the four search engines were then combined by iProphet to merge the different, overlapping subsets of identified sample peptides to provide a more accurate and conservative representation of the identified peptides [160]. Finally, we used PTMProphet to verify the phosphorylation site localization. Raw mass spectrometry data is deposited in the PeptideAtlas and can be accessed at http://www.peptideatlas.org/PASS/.

Analysis of Public Mtb Mass Spectrometry Data Sets

Data published by Schubert *et al.* [150] and Kelkar *et al.* [166] were available through the public data repository PeptideAtlas (www.peptideatlas.org) while the data set by Huang and coworkers was obtained through Peptidome (www.ncbi.nlm.nih.gov/peptidome). Data in mzData file format were converted to mzXML and searched against the *Mtb* proteome using Comet and X!Tandem. Search results were further processed using the Trans-Proteomic Pipeline (TPP, v4.6 rev1) including PeptideProphet [159] and iProphet [160]. The search parameters included a static modification of +57.021464 to account for carbamidomethylated cysteines and variable modifications of +15.9949 for oxidized methionines, +79.966331 for phosphorylated serines, threonines and tyrosines, +42.0106 for *N*-terminal acetylation. X!Tandem enables by default variable modifications to account for additional *N*-terminal modifications including +17.0265 at

glutamine, -18.0106 at glutamic acid and -17.0265 at cysteine. The parent mass error was set to ± 20 ppm, missed cleavages were set to two and semi-cleavages allowed. For X!Tandem the parent monoisotopic mass isotope error was enabled. PeptideProphet was run with accurate mass binning and parametric model. PeptideProphet results from Comet and X!Tandem were combined using iProphet. The error rate for PSMs was estimated by iProphet for each experiment and only peptides with a minimum probability of 0.9 were considered. One peptide identification was discarded despite a high iProphet probability as tyrosine phosphorylation was not sufficient supported by the obtained fragmentation pattern.

Supporting Figures

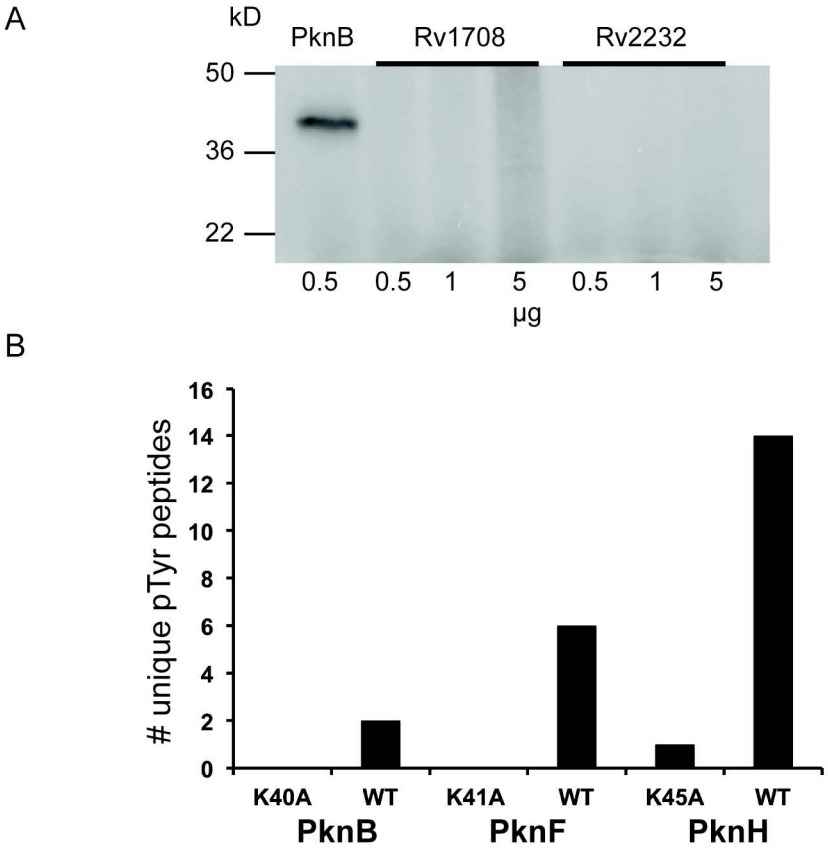
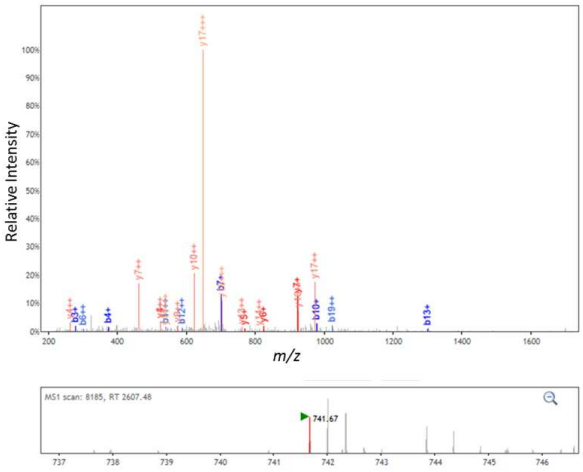


Figure 4-S1. Rv1708 and Rv2232 have no detectable kinase activity. (A) Autoradiograph of PknB, Rv1708, and Rv2232 autophosphorylating activity shows phosphotransfer by PknB, but not Rv1708 and Rv2232, even at 10-fold higher protein concentration. (B) PknB, F, and H have intrinsic Tyr autophosphorylating activity. Unique pTyr peptides from recombinantly expressed wild type (WT) and kinase dead (K/A) mutants were determined by LC-MS/MS. Tyr phosphorylation on wild-type, but not on kinase dead mutants shows that pTyr is not a result of *E. coli* phosphorylation, but of intrinsic autophosphorylation.

R.GGYPPETGGYPPQPG**pY**(1.000)PRPR.H
Rv0020c

endogenous *Mtb* H37Rv

iProphet probability 0.999581, PTMProphet score 1.000
MH^+ 2223.0019, precursor charge 3+, scan 8187, exp. *m/z* 741.6711, mass accuracy -1.4832 ppm

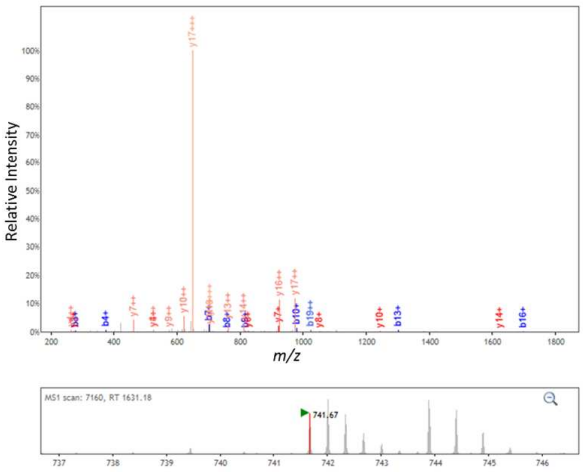


b+	b2+	#	Seq	#	y+	y2+	y3+
58.0287	29.5180	1	G	20			
115.0502	58.0287	2	G	19	2165.9805	1083.4939	722.6650
278.1135	139.5604	3	Y	18	2108.9590	1054.9831	703.6579
375.1663	188.0868	4	P	17	1945.8957	973.4515	649.3034
472.2191	236.6132	5	P	16	1848.8429	924.9251	616.9525
601.2617	301.1345	6	E	15	1751.7902	876.3987	584.6016
702.3093	351.6583	7	T	14	1622.7476	811.8774	541.5874
759.3308	380.1690	8	G	13	1521.6999	761.3536	507.9048
816.3523	408.6798	9	G	12	1464.6784	732.8428	488.8977
979.4156	490.2114	10	Y	11	1407.6570	704.3321	469.8905
1076.4684	538.7378	11	P	10	1244.5936	622.8004	415.5361
1173.5211	587.2642	12	P	9	1147.5409	574.2741	383.1851
1301.5797	651.2935	13	Q	8	1050.4881	525.7477	350.8342
1398.6325	699.8199	14	P	7	922.4295	461.7184	308.1480
1455.6539	728.3306	15	G	6	825.3768	413.1920	275.7971
1698.6836	849.8454	16	Y	5	768.3553	384.6813	256.7899
1795.7364	898.3718	17	P	4	525.3256	263.1664	175.7800
1951.8375	976.4224	18	R	3	428.2728	214.6401	143.4291
2048.8903	1024.9488	19	P	2	272.1717	136.5895	91.3954
		20	R	1	175.1190	88.0631	59.0445

R.GGYPPETGGYPPQPG**pY**(1.000)PRPR.H
Rv0020c

synthetic peptide

iProphet probability 0.999837, PTMProphet score 1.000
MH^+ 2223.0019, precursor charge 3+, scan 7161, exp. *m/z* 741.6713, mass accuracy -1.2134 ppm

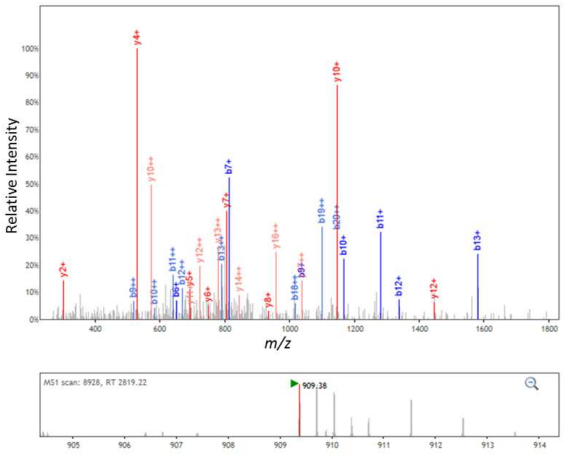


b+	b2+	#	Seq	#	y+	y2+	y3+
58.0287	29.5180	1	G	20			
115.0502	58.0287	2	G	19	2165.9805	1083.4939	722.6650
278.1135	139.5604	3	Y	18	2108.9590	1054.9831	703.6579
375.1663	188.0868	4	P	17	1945.8957	973.4515	649.3034
472.2191	236.6132	5	P	16	1848.8429	924.9251	616.9525
601.2617	301.1345	6	E	15	1751.7902	876.3987	584.6016
702.3093	351.6583	7	T	14	1622.7476	811.8774	541.5874
759.3308	380.1690	8	G	13	1521.6999	761.3536	507.9048
816.3523	408.6798	9	G	12	1464.6784	732.8428	488.8977
979.4156	490.2114	10	Y	11	1407.6570	704.3321	469.8905
1076.4684	538.7378	11	P	10	1244.5936	622.8004	415.5361
1173.5211	587.2642	12	P	9	1147.5409	574.2741	383.1851
1301.5797	651.2935	13	Q	8	1050.4881	525.7477	350.8342
1398.6325	699.8199	14	P	7	922.4295	461.7184	308.1480
1455.6539	728.3306	15	G	6	825.3768	413.1920	275.7971
1698.6836	849.8454	16	Y	5	768.3553	384.6813	256.7899
1795.7364	898.3718	17	P	4	525.3256	263.1664	175.7800
1951.8375	976.4224	18	R	3	428.2728	214.6401	143.4291
2048.8903	1024.9488	19	P	2	272.1717	136.5895	91.3954
		20	R	1	175.1190	88.0631	59.0445

R.HPDQGDYPEQIG**pY**(1.000)PDQGGYPEQR.G
Rv0020c

endogenous *Mtb* H37Rv

iProphet probability 0.995576, PTMProphet score 1.000
MH^+ 2726.1155, precursor charge 3+, scan 8929, exp. *m/z* 909.3779, mass accuracy 1.3195 ppm

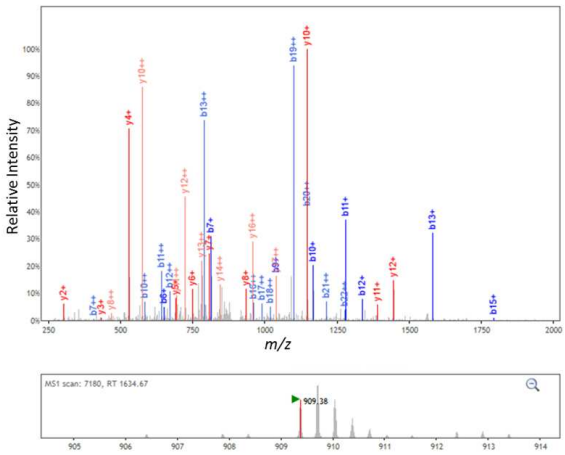


b+	b2+	#	Seq	#	y+	y2+
138.0662	69.5367	1	H	23		
235.1190	118.0631	2	P	22	2589.0566	1295.0319
350.1459	175.5766	3	D	21	2492.0038	1246.5056
478.2045	239.6059	4	Q	20	2376.9769	1188.9921
535.2259	268.1166	5	G	19	2248.9183	1124.9628
650.2529	325.6301	6	D	18	2191.8969	1096.4521
813.3162	407.1617	7	Y	17	2076.8699	1038.9386
910.3690	455.6881	8	P	16	1913.8066	957.4069
1039.4116	520.2094	9	E	15	1816.7538	908.8806
1167.4701	584.2387	10	Q	14	1687.7112	844.3593
1280.5542	640.7807	11	I	13	1559.6527	780.3300
1337.5757	669.2915	12	G	12	1446.5686	723.7879
1580.6054	790.8063	13	Y	11	1389.5471	695.2772
1677.6581	839.3327	14	P	10	1146.5174	573.7624
1792.6851	896.8462	15	D	9	1049.4647	525.2360
1920.7437	960.8755	16	Q	8	934.4377	467.7225
1977.7651	989.3862	17	G	7	806.3791	403.6932
2034.7866	1017.8969	18	G	6	749.3577	375.1825
2197.8499	1099.4286	19	Y	5	692.3362	346.6717
2294.9027	1147.9550	20	P	4	529.2729	265.1401
2423.9453	1212.4763	21	E	3	432.2201	216.6137
2552.0038	1276.5056	22	Q	2	303.1775	152.0924
		23	R	1	175.1190	88.0631

R.HPDQGDYPEQIG**pY**(1.000)PDQGGYPEQR.G
Rv0020c

synthetic peptide

iProphet probability 0.982579, PTMProphet score 1.000
MH^+ 2726.1155, precursor charge 3+, scan 7187, exp. *m/z* 909.3777, mass accuracy 1.0996 ppm

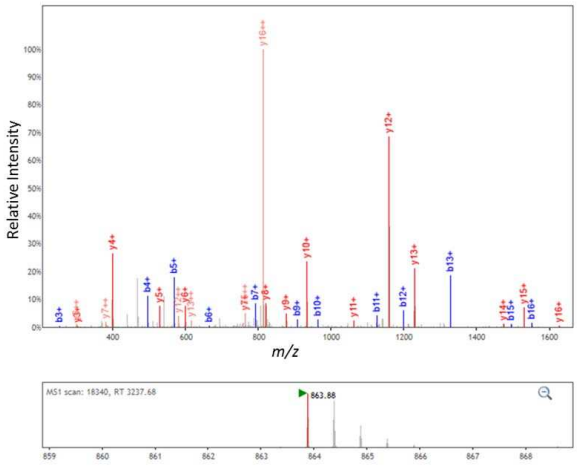


b+	b2+	#	Seq	#	y+	y2+
138.0662	69.5367	1	H	23		
235.1190	118.0631	2	P	22	2589.0566	1295.0319
350.1459	175.5766	3	D	21	2492.0038	1246.5056
478.2045	239.6059	4	Q	20	2376.9769	1188.9921
535.2259	268.1166	5	G	19	2248.9183	1124.9628
650.2529	325.6301	6	D	18	2191.8969	1096.4521
813.3162	407.1617	7	Y	17	2076.8699	1038.9386
910.3690	455.6881	8	P	16	1913.8066	957.4069
1039.4116	520.2094	9	E	15	1816.7538	908.8806
1167.4701	584.2387	10	Q	14	1687.7112	844.3593
1280.5542	640.7807	11	I	13	1559.6527	780.3300
1337.5757	669.2915	12	G	12	1446.5686	723.7879
1580.6054	790.8063	13	Y	11	1389.5471	695.2772
1677.6581	839.3327	14	P	10	1146.5174	573.7624
1792.6851	896.8462	15	D	9	1049.4647	525.2360
1920.7437	960.8755	16	Q	8	934.4377	467.7225
1977.7651	989.3862	17	G	7	806.3791	403.6932
2034.7866	1017.8969	18	G	6	749.3577	375.1825
2197.8499	1099.4286	19	Y	5	692.3362	346.6717
2294.9027	1147.9550	20	P	4	529.2729	265.1401
2423.9453	1212.4763	21	E	3	432.2201	216.6137
2552.0038	1276.5056	22	Q	2	303.1775	152.0924
		23	R	1	175.1190	88.0631

R.VPG**pY**(1.000)APQGGGYAEPAGR.D
Rv0020c

endogenous *Mtb* H37Rv

iProphet probability 0.999786, PTMProphet score 1.000
MH^+ 1726.7585, precursor charge 2+, scan 18348, exp. *m/z* 863.8823, mass accuracy -0.6945 ppm

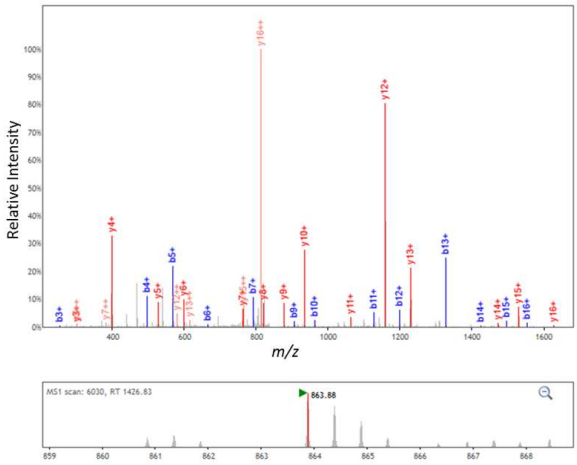


b+	#	Seq	#	y+	y2+
100.0757	1	V	17		
197.1285	2	P	16	1627.6901	814.3487
254.1499	3	G	15	1530.6373	765.8223
497.1796	4	Y	14	1473.6158	737.3116
568.2167	5	A	13	1230.5862	615.7967
665.2695	6	P	12	1159.5491	580.2782
793.3280	7	Q	11	1062.4963	531.7518
850.3495	8	G	10	934.4377	467.7225
907.3710	9	G	9	877.4163	439.2118
964.3924	10	G	8	820.3948	410.7010
1127.4558	11	Y	7	763.3733	382.1903
1198.4929	12	A	6	600.3100	300.6586
1327.5355	13	E	5	529.2729	265.1401
1424.5882	14	P	4	400.2303	200.6188
1495.6253	15	A	3	303.1775	152.0924
1552.6468	16	G	2	232.1404	116.5738
	17	R	1	175.1190	88.0631

R.VPG**pY**(1.000)APQGGGYAEPAGR.D
Rv0020c

synthetic peptide

iProphet probability 0.999997, PTMProphet score 1.000
MH^+ 1726.7585, precursor charge 2+, scan 6032, exp. *m/z* 863.8838, mass accuracy 1.0418 ppm

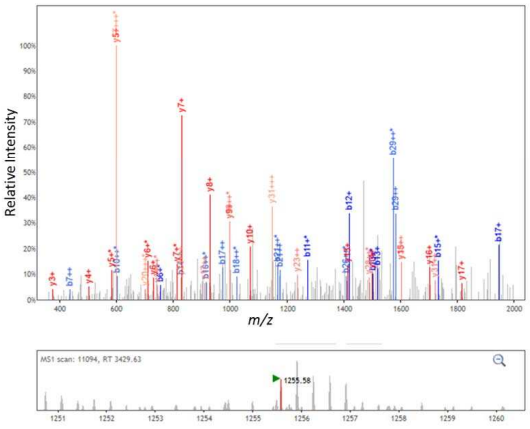


b+	#	Seq	#	y+	y2+
100.0757	1	V	17		
197.1285	2	P	16	1627.6901	814.3487
254.1499	3	G	15	1530.6373	765.8223
497.1796	4	Y	14	1473.6159	737.3116
568.2167	5	A	13	1230.5862	615.7967
665.2695	6	P	12	1159.5491	580.2782
793.3281	7	Q	11	1062.4963	531.7518
850.3495	8	G	10	934.4377	467.7225
907.3710	9	G	9	877.4163	439.2118
964.3925	10	G	8	820.3948	410.7010
1127.4558	11	Y	7	763.3733	382.1903
1198.4929	12	A	6	600.3100	300.6586
1327.5355	13	E	5	529.2729	265.1401
1424.5883	14	P	4	400.2303	200.6188
1495.6254	15	A	3	303.1775	152.0924
1552.6468	16	G	2	232.1404	116.5738
	17	R	1	175.1190	88.0631

R.TES**pY**(0.614)PLVPHDAETETVVITTSDNDAAVTQPEAQR.E
Rv0361

endogenous *Mtb* H37Rv

iProphet probability 0.999819, PTMProphet score 0.614
MH^+ 3764.7175, precursor charge 3+, scan 11096,
exp. *m/z* 1255.5789, mass accuracy 1.1946 ppm

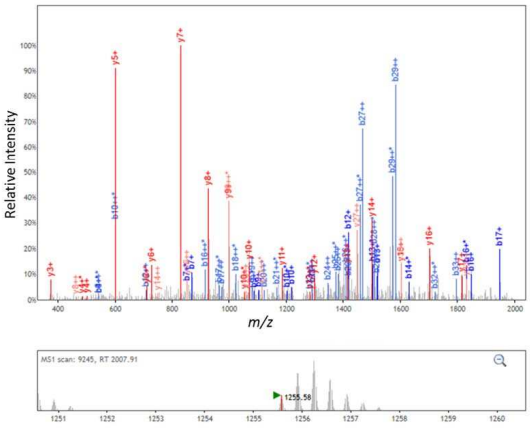


b+	b2+	#	Seq	#	y+	y2+	y3+
102.0550	51.5311	1	T	34			
231.0975	116.0524	2	E	33	3663.6698	1832.3386	1221.8948
318.1296	159.5684	3	S	32	3534.6272	1767.8173	1178.8806
561.1593	281.0833	4	Y	31	3447.5952	1724.3012	1149.8699
658.2120	329.6097	5	P	30	3204.5655	1602.7864	1068.8600
771.2961	386.1517	6	L	29	3107.5127	1554.2600	1036.5091
870.3645	435.6859	7	V	28	2994.4287	1497.7180	998.8144
967.4173	484.2123	8	P	27	2895.3603	1448.1838	965.7916
1104.4762	552.7417	9	H	26	2798.3075	1399.6574	933.4407
1219.5031	610.2552	10	D	25	2661.2486	1331.1279	887.7544
1290.5403	645.7738	11	A	24	2546.2216	1273.6145	849.4121
1419.5828	710.2951	12	E	23	2475.1845	1238.0959	825.7330
1520.6305	760.8189	13	T	22	2346.1419	1173.5746	782.7188
1649.6731	825.3402	14	E	21	2245.0943	1123.0508	749.0363
1750.7208	875.8640	15	T	20	2116.0517	1058.5295	706.0221
1849.7892	925.3982	16	V	19	2015.0040	1008.0056	672.3395
1948.8576	974.9324	17	V	18	1915.9356	958.4714	639.3167
2061.9417	1031.4745	18	I	17	1816.8672	908.9372	606.2939
2162.9894	1081.9983	19	T	16	1703.7831	852.3952	568.5992
2264.0370	1132.5222	20	T	15	1602.7354	801.8713	534.9167
2351.0691	1176.0382	21	S	14	1501.6877	751.3475	501.2341
2466.0960	1233.5516	22	D	13	1414.6557	707.8315	472.2234
2580.1389	1290.5731	23	N	12	1299.6288	650.3180	433.8811
2695.1659	1348.0866	24	D	11	1185.5858	593.2966	395.8668
2766.2030	1383.6051	25	A	10	1070.5589	535.7831	357.5245
2837.2401	1419.1237	26	A	9	999.5218	500.2645	333.8454
2936.3085	1468.6579	27	V	8	928.4847	464.7460	310.1664
3037.3562	1519.1817	28	T	7	829.4163	415.2118	277.1436
3165.4148	1583.2110	29	Q	6	728.3686	364.6879	243.4610
3262.4675	1631.7374	30	P	5	600.3100	300.6586	200.7749
3391.5101	1696.2587	31	E	4	503.2572	252.1323	168.4239
3462.5473	1731.7773	32	A	3	374.2146	187.6110	125.4097
3590.6058	1795.8066	33	Q	2	303.1775	152.0924	101.7307
		34	R	1	175.1190	88.0631	59.0445

R.TES**pY**(0.368)PLVPHDAETETVVITTSDNDAAVTQPEAQR.E
Rv0361

synthetic peptide

iProphet probability 0.960535, PTMProphet score 0.368
MH^+ 3764.7175, precursor charge 3+, scan 9248,
exp. *m/z* 1255.5792, mass accuracy 1.4336 ppm

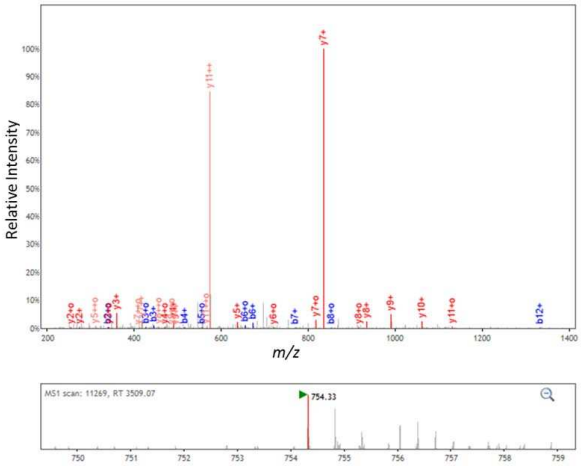


b+	b2+	#	Seq	#	y+	y2+
102.0550	51.5311	1	T	34		
231.0975	116.0524	2	E	33	3663.6698	1832.3386
318.1296	159.5684	3	S	32	3534.6272	1767.8173
561.1593	281.0833	4	Y	31	3447.5952	1724.3012
658.2120	329.6097	5	P	30	3204.5655	1602.7864
771.2961	386.1517	6	L	29	3107.5127	1554.2600
870.3645	435.6859	7	V	28	2994.4287	1497.7180
967.4173	484.2123	8	P	27	2895.3603	1448.1838
1104.4762	552.7417	9	H	26	2798.3075	1399.6574
1219.5031	610.2552	10	D	25	2661.2486	1331.1279
1290.5403	645.7738	11	A	24	2546.2216	1273.6145
1419.5828	710.2951	12	E	23	2475.1845	1238.0959
1520.6305	760.8189	13	T	22	2346.1419	1173.5746
1649.6731	825.3402	14	E	21	2245.0943	1123.0508
1750.7208	875.8640	15	T	20	2116.0517	1058.5295
1849.7892	925.3982	16	V	19	2015.0040	1008.0056
1948.8576	974.9324	17	V	18	1915.9356	958.4714
2061.9417	1031.4745	18	I	17	1816.8672	908.9372
2162.9894	1081.9983	19	T	16	1703.7831	852.3952
2264.0370	1132.5222	20	T	15	1602.7354	801.8713
2351.0691	1176.0382	21	S	14	1501.6877	751.3475
2466.0960	1233.5516	22	D	13	1414.6557	707.8315
2580.1389	1290.5731	23	N	12	1299.6288	650.3180
2695.1659	1348.0866	24	D	11	1185.5858	593.2966
2766.2030	1383.6051	25	A	10	1070.5589	535.7831
2837.2401	1419.1237	26	A	9	999.5218	500.2645
2936.3085	1468.6579	27	V	8	928.4847	464.7460
3037.3562	1519.1817	28	T	7	829.4163	415.2118
3165.4148	1583.2110	29	Q	6	728.3686	364.6879
3262.4675	1631.7374	30	P	5	600.3100	300.6586
3391.5101	1696.2587	31	E	4	503.2572	252.1323
3462.5473	1731.7773	32	A	3	374.2146	187.6110
3590.6058	1795.8066	33	Q	2	303.1775	152.0924
		34	R	1	175.1190	88.0631

R.**pY**(1.000)DSAGVPTFESVR.E
Rv1638A

endogenous *Mtb* H37Rv

iProphet probability 0.991416, PTMProphet score 1.000
MH^+ 1507.6465, precursor charge 2+, scan 11273, exp. *m/z* 754.3279, mass accuracy -0.6945 ppm

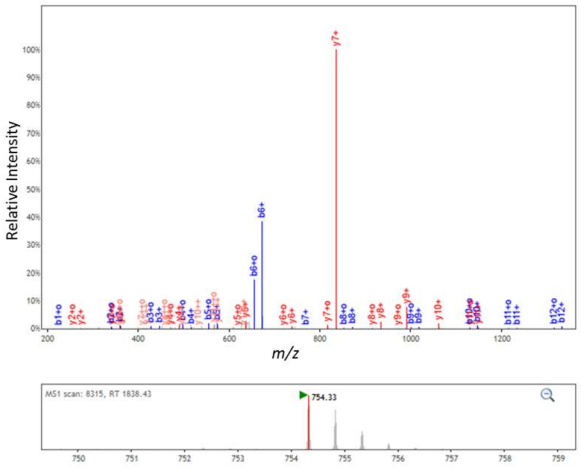


b+	#	Seq	#	y+	y2+
244.0370	1	Y	13		
359.0639	2	D	12	1264.6168	632.8120
446.0959	3	S	11	1149.5899	575.2986
517.1331	4	A	10	1062.5578	531.7826
574.1545	5	G	9	991.5207	496.2640
673.2229	6	V	8	934.4993	467.7533
770.2757	7	P	7	835.4308	418.2191
871.3234	8	T	6	738.3781	369.6927
1018.3918	9	F	5	637.3304	319.1688
1147.4344	10	E	4	490.2620	245.6346
1234.4664	11	S	3	361.2194	181.1133
1333.5348	12	V	2	274.1874	137.5973
	13	R	1	175.1190	88.0631

R.**pY**(1.000)DSAGVPTFESVR.E
Rv1638A

synthetic peptide

iProphet probability 0.991416, PTMProphet score 1.000
MH^+ 1507.6465, precursor charge 2+, scan 8319, exp. *m/z* 754.3273, mass accuracy 0.53027 ppm

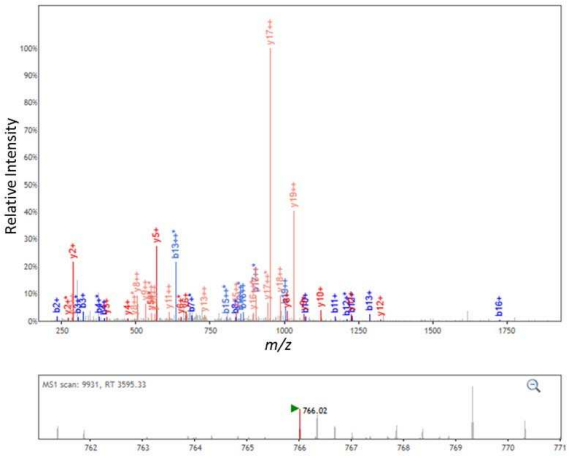


b+	#	Seq	#	y+	y2+
244.0370	1	Y	13		
359.0639	2	D	12	1264.6168	632.8120
446.0959	3	S	11	1149.5899	575.2986
517.1331	4	A	10	1062.5578	531.7826
574.1545	5	G	9	991.5207	496.2640
673.2229	6	V	8	934.4993	467.7533
770.2757	7	P	7	835.4308	418.2191
871.3234	8	T	6	738.3781	369.6927
1018.3918	9	F	5	637.3304	319.1688
1147.4344	10	E	4	490.2620	245.6346
1234.4664	11	S	3	361.2194	181.1133
1333.5348	12	V	2	274.1874	137.5973
	13	R	1	175.1190	88.0631

R.AYSAPESEHVTGGP**pY**(1.000)VPADLR.L
Rv2198c

endogenous *Mtb* H37Rv

iProphet probability 0.999999, PTMProphet score 1.000
MH^+ 2296.0282, precursor charge 3+, scan 9935, exp. *m/z* 766.0157, mass accuracy 1.8276 ppm

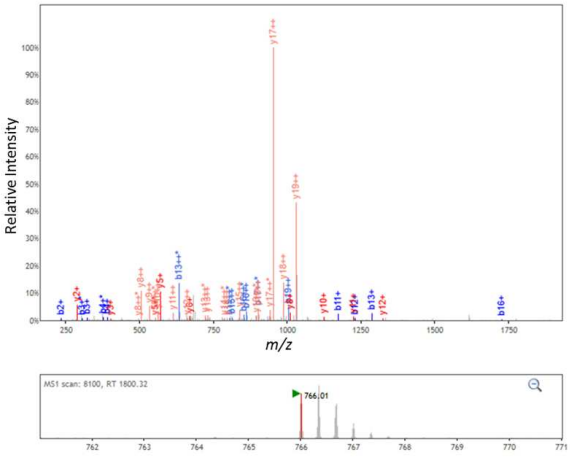


b+	b2+	#	Seq	#	y+	y2+
72.0444	36.5258	1	A	21		
235.1077	118.0575	2	Y	20	2224.9911	1112.9992
322.1397	161.5735	3	S	19	2061.9278	1031.4675
393.1769	197.0921	4	A	18	1974.8957	987.9515
490.2296	245.6185	5	P	17	1903.8586	952.4330
619.2722	310.1397	6	E	16	1806.8059	903.9066
706.3042	353.6558	7	S	15	1677.7633	839.3853
835.3468	418.1771	8	E	14	1590.7312	795.8693
972.4058	486.7065	9	H	13	1461.6886	731.3480
1071.4742	536.2407	10	V	12	1324.6297	662.8185
1172.5218	586.7646	11	T	11	1225.5613	613.2843
1229.5433	615.2753	12	G	10	1124.5136	562.7605
1286.5648	643.7860	13	G	9	1067.4922	534.2497
1383.6175	692.3124	14	P	8	1010.4707	505.7390
1626.6472	813.8273	15	Y	7	913.4180	457.2126
1725.7156	863.3615	16	V	6	670.3883	335.6978
1822.7684	911.8878	17	P	5	571.3198	286.1636
1893.8055	947.4064	18	A	4	474.2671	237.6372
2008.8325	1004.9199	19	D	3	403.2300	202.1186
2121.9165	1061.4619	20	L	2	288.2030	144.6051
		21	R	1	175.1190	88.0631

R.AYSAPESEHVTGGP**pY**(1.000)VPADLR.L
Rv2198c

synthetic peptide

iProphet probability 0.999999, PTMProphet score 1.000
MH^+ 2296.0282, precursor charge 3+, scan 8101, exp. *m/z* 766.0143, mass accuracy 0 ppm

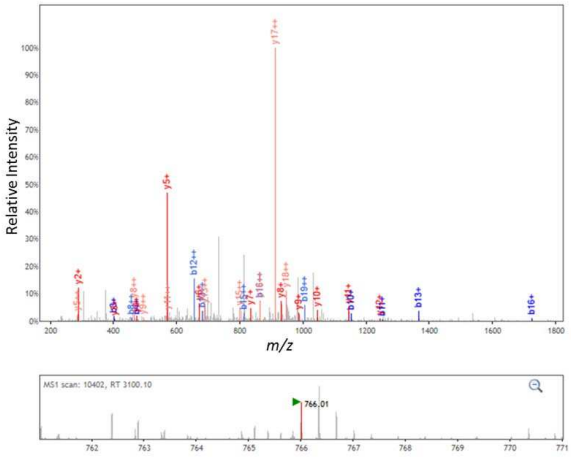


b+	b2+	#	Seq	#	y+	y2+
72.0444	36.5258	1	A	21		
235.1077	118.0575	2	Y	20	2224.9911	1112.9992
322.1397	161.5735	3	S	19	2061.9278	1031.4675
393.1769	197.0921	4	A	18	1974.8957	987.9515
490.2296	245.6185	5	P	17	1903.8586	952.4330
619.2722	310.1397	6	E	16	1806.8059	903.9066
706.3042	353.6558	7	S	15	1677.7633	839.3853
835.3468	418.1771	8	E	14	1590.7312	795.8693
972.4058	486.7065	9	H	13	1461.6886	731.3480
1071.4742	536.2407	10	V	12	1324.6297	662.8185
1172.5218	586.7646	11	T	11	1225.5613	613.2843
1229.5433	615.2753	12	G	10	1124.5136	562.7605
1286.5648	643.7860	13	G	9	1067.4922	534.2497
1383.6175	692.3124	14	P	8	1010.4707	505.7390
1626.6472	813.8273	15	Y	7	913.4180	457.2126
1725.7156	863.3615	16	V	6	670.3883	335.6978
1822.7684	911.8878	17	P	5	571.3198	286.1636
1893.8055	947.4064	18	A	4	474.2671	237.6372
2008.8325	1004.9199	19	D	3	403.2300	202.1186
2121.9165	1061.4619	20	L	2	288.2030	144.6051
		21	R	1	175.1190	88.0631

R.A**pY**(0.527)SAPESEHVTGGPYVPADLR.L
Rv2198c

endogenous *Mtb* H37Rv

iProphet probability 0.999472, PTMProphet score 0.527
MH^+ 2296.0282, precursor charge 3+, scan 10403, exp. *m/z* 766.0136, mass accuracy -0.9138 ppm

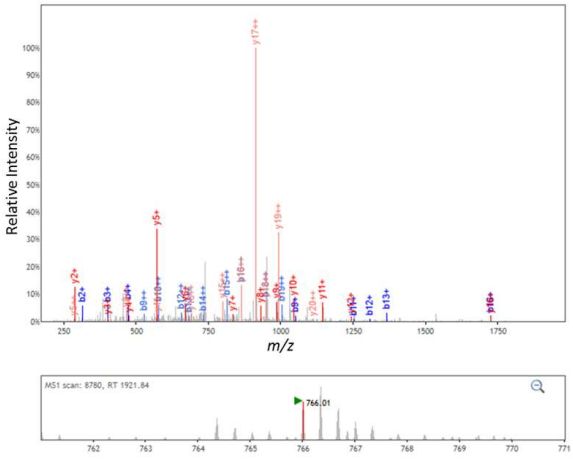


b+	b2+	#	Seq	#	y+	y2+
72.0444	36.5258	1	A	21		
315.0741	158.0407	2	Y	20	2224.9911	1112.9992
402.1061	201.5567	3	S	19	1981.9614	991.4843
473.1432	237.0753	4	A	18	1894.9294	947.9683
570.1960	285.6016	5	P	17	1823.8923	912.4498
699.2386	350.1229	6	E	16	1726.8395	863.9234
786.2706	393.6389	7	S	15	1597.7969	799.4021
915.3132	458.1602	8	E	14	1510.7649	755.8861
1052.3721	526.6897	9	H	13	1381.7223	691.3648
1151.4405	576.2239	10	V	12	1244.6634	622.8353
1252.4882	626.7477	11	T	11	1145.5949	573.3011
1309.5097	655.2585	12	G	10	1044.5473	522.7773
1366.5311	683.7692	13	G	9	987.5258	494.2665
1463.5839	732.2956	14	P	8	930.5043	465.7558
1626.6472	813.8273	15	Y	7	833.4516	417.2294
1725.7156	863.3615	16	V	6	670.3883	335.6978
1822.7684	911.8878	17	P	5	571.3198	286.1636
1893.8055	947.4064	18	A	4	474.2671	237.6372
2008.8325	1004.9199	19	D	3	403.2300	202.1186
2121.9165	1061.4619	20	L	2	288.2030	144.6051
		21	R	1	175.1190	88.0631

R.A**pY**(1.000)SAPESEHVTGGPYVPADLR.L
Rv2198c

synthetic peptide

iProphet probability 0.999999, PTMProphet score 1.000
MH^+ 2296.0282, precursor charge 3+, scan 8782, exp. *m/z* 766.0145, mass accuracy 0.2611 ppm



b+	b2+	#	Seq	#	y+	y2+
72.0444	36.5258	1	A	21		
315.0741	158.0407	2	Y	20	2224.9911	1112.9992
402.1061	201.5567	3	S	19	1981.9614	991.4843
473.1432	237.0753	4	A	18	1894.9294	947.9683
570.1960	285.6016	5	P	17	1823.8923	912.4498
699.2386	350.1229	6	E	16	1726.8395	863.9234
786.2706	393.6389	7	S	15	1597.7969	799.4021
915.3132	458.1602	8	E	14	1510.7649	755.8861
1052.3721	526.6897	9	H	13	1381.7223	691.3648
1151.4405	576.2239	10	V	12	1244.6634	622.8353
1252.4882	626.7477	11	T	11	1145.5949	573.3011
1309.5097	655.2585	12	G	10	1044.5473	522.7773
1366.5311	683.7692	13	G	9	987.5258	494.2665
1463.5839	732.2956	14	P	8	930.5043	465.7558
1626.6472	813.8273	15	Y	7	833.4516	417.2294
1725.7156	863.3615	16	V	6	670.3883	335.6978
1822.7684	911.8878	17	P	5	571.3198	286.1636
1893.8055	947.4064	18	A	4	474.2671	237.6372
2008.8325	1004.9199	19	D	3	403.2300	202.1186
2121.9165	1061.4619	20	L	2	288.2030	144.6051
		21	R	1	175.1190	88.0631

Figure
identi
simila
Spectr
Peaks
blue, corresponding fragment masses are listed in the table, the phosphotyrosine residue is highlighted in yellow. Neutral losses are indicated by 'o' for H_2O and '*' for NH_3 and only displayed in spectra with sufficient room for annotation. The MS1 scan is displayed below the MS/MS spectrum.

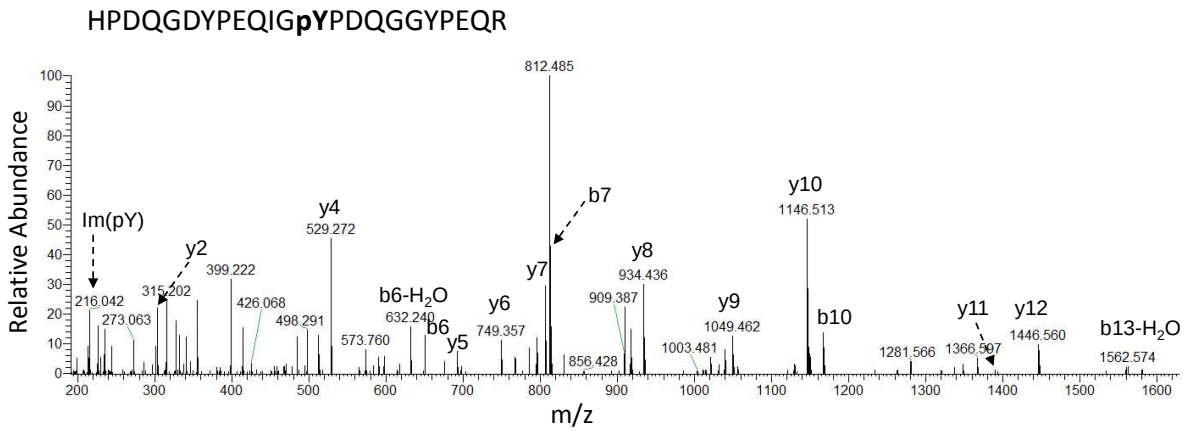


Figure 4-S3. Phosphotyrosine specific reporter ion. MS/MS spectrum of tyrosine-phosphorylated peptide HPDQGDYPEQIGpYPDQGGYPEQR identified in *Mtb* lysate on a Q Exactive (precursor mass accuracy 2.09 ppm). The position of the phosphorylated tyrosine is indicated as pY. Fragment b and y ions are annotated including the phosphotyrosine specific immonium ion Im(pY) at *m/z* 216.042.

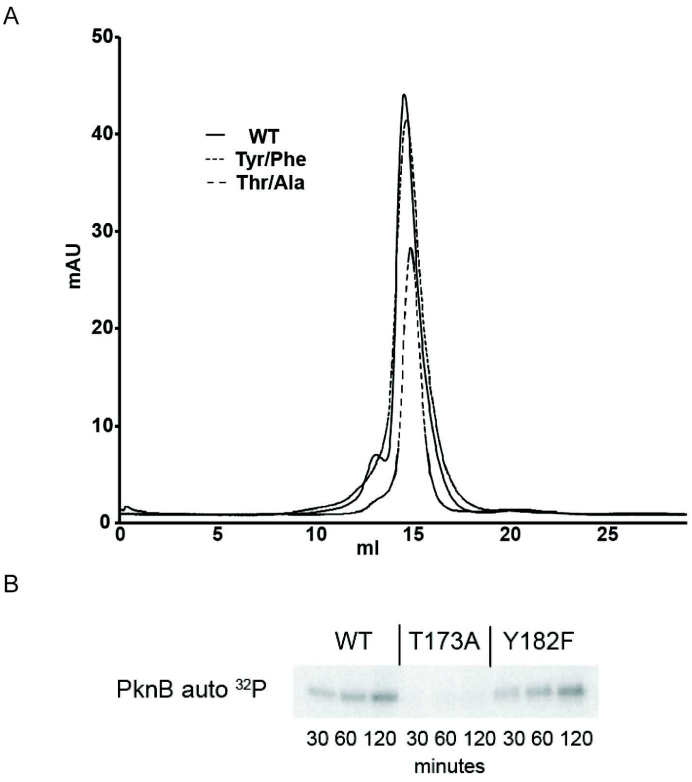


Figure 4-S4. PknB mutants are properly folded and autophosphorylation is not affected by Tyr phosphosites loss. (A) Size exclusion chromatography profiles show that the Tyr/Phe and Thr/Ala mutants elute at the same volume as wild type. (B) Autoradiograph of ^{32}P incorporation shows similar autophosphorylating activity of the Tyr/Phe mutant and wild type.

Table S1. Phosphotyrosine peptide identifications in recombinant *Mtb* STPKs

STPK[a]	pY peptide sequence with PTMProphetscore[b]	iProphet[c]	pY site[d]	Sequence coverage[e]
Rv0014c	***Transmembrane serine/threonine-protein kinase PknB***			
pknB	R.**pY**(1.000)[f]ELGEILGFGGMSEVHLAR.D	0.970064	pY11	47.6%
	R.DPSF**pY**(1.000)LR.F	0.999956	pY53	
	A.IVAV**pY**(0.679)DTGEAETPAGPLPYIVMEYVDGVTLR.D	0.968486	pY75	
	A.V**pY**(0.546)DTGEAETPAGPLPYIVMEYVDGVTLR.D	0.926652	pY75	
	R.EAQNAAALNHPAIVAVYDTGEAETPAGPLP**pY**(1.000)IVME**pY**(1.000)V.D	0.997037	pY89, pY94	
	Y.DTGEAETPAGPLP**pY**(1.000)VMEYVDGVTLR.D	0.994719	pY89	
	R.AIADSGNSVTQTAAVIGTAQ**pY**(1.000)LSPEQAR.G	0.999999	pY182	
	R.AIADSGN**pS**(0.523)VTQTAAVIGTAQ**pY**(0.517)LSPEQAR.G	0.999999	pY182	
	R.AIADSGNSVTQ**pT**(0.542)AAVIGTAQ**pY**(0.440)LSPEQAR.G	0.999999	pY182	
	R.AIAD**pS**(0.840)GN**pS**(0.573)VTQTAAVIGTAQ**pY**(0.613)LSPEQAR.G	0.999999	pY182	
	R.AIADSGN**pS**(0.649)V**pT**(0.730)QTAAVIGTAQ**pY**(0.605)LSPEQAR	0.999997	pY182	
	R.AIADSGNSV**pT**(0.718)Q**pT**(0.741)AAVIGTAQ**pY**(0.615)LSPEQAR.G	0.999999	pY182	
	R.AIADSGNSV**pT**(0.728)QTAAVIG**pT**(0.839)AQ**pY**(0.788)LSPEQAR.G	0.999997	pY182	
	R.AIADSGNSVTQ**pT**(0.817)AAVIG**pT**(0.750)AQ**pY**(0.817)LSPEQAR.G	0.999999	pY182	
	R.AIADSGN**pS**(0.788)V**pT**(0.872)Q**pT**(0.832)AAVIGTAQ**pY**(0.613)LSPEQAR.G	0.990566	pY182	
	R.SDV**pY**(0.444)SLGCVLYEVLTGEPPFTGDSPVSVAYQHVR.E	0.999532	pY200	
	Y.SLGCVL**pY**(0.998)EVLTGEPPFTGDSPVSVAYQHVR.E	0.985982	pY207	
	G.CVL**pY**(0.687)EVLTGEPPFTGDSPVSVAYQHVR.E	0.922437	pY207	
	R.SDVYSLGCVLYEVLTGEPPFTGDSPVSVA**pY**(0.993)QHVR.E	0.999707	pY226	
	Y.SLGCVLYEVLTGEPPFTGDSPVSVA**pY**(1.000)QHVR.E	0.994889	pY226	
	C.VLYEVLTGEPPFTGDSPVSVA**pY**(0.687)QHVR.E	0.907812	pY226	
	L.TGEPPFTGDSPVSVA**pY**(1.000)QHVR.E	0.965062	pY226	
	T.GEPPFTGDSPVSVA**pY**(1.000)QHVR.E	0.999987	pY226	
Rv0931c	***Transmembrane serine/threonine-protein kinase PknD***			
pknD	R.GGMGEV**pY**(1.000)EAEDTR.K	0.999998	pY30	53.6 %
	G.GMGEV**pY**(1.000)EAEDTR.K	0.994992	pY30	
	G.GMGEV**pY**(1.000)EAED**pT**(1.000)R.K	0.965579	pY30	
	K.LISPQ**pY**(0.477)SDNAVFR.A	0.999996	pY50	
	R.LTEPHIVPIHD**pY**(0.998)GEINGQFFVEMR.M	0.999996	pY81	
	R.AA**pS**(0.915)DPGLTQTG**pT**(0.890)AVGT**pY**(0.989)NYMAPER.F	0.999999	pY178	
	R.AASDPGL**pT**(0.898)Q**pT**(0.860)GTAVGT**pY**(0.735)NYMAPER.F	0.999999	pY178	
	R.AASDPGLTQ**pT**(0.698)G**pT**(0.698)AVGT**pY**(0.821)NYMAPER.F	0.999999	pY178	
	R.AASDPGLTQTGTAVG**pT**(0.465)**pY**(0.465)NYMAPER.F	0.999999	pY178	
	R.AASDPGLTQ**pT**(0.946)G**pT**(0.735)AVGTYN**pY**(1.000)MAPER.F	0.999999	pY180	
	R.ADI**pY**(1.000)ALACVLGECLTGAPPYR.A	0.999951	pY198	
Rv1743	***Probable transmembrane serine/threonine-protein kinase PknE***			
pknE	R.EG**pT**(1.000)QFGP**pY**(1.000)R.L	0.999347	pY16	47.3 %
	R.GGMGDV**pY**(1.000)EAEDTVR.E	0.999999	pY31	

	R.GGM$_{ox}$GDV**pY**(1.000)EAEDTVR.E	0.999999	pY31	
	R.ADIYALTCVLYECLTGSPP**pY**(0.989)QGDQLSVMGAHINQAIPR.P	0.992325	pY215	
	R.ADIYALTCVL**pY**(0.807)ECLTGSPPYQGDQLSVMGAHINQAIPR.P	0.934447	pY206	
	K.L**pT**(0.960)QLGNTVGTL**pY**(0.998)YMAPER.F	0.995084	pY180	
	R.**pY**(0.773)VTCGDLSAAAHAALATADQDR.A	0.999999	pY261	
Rv1746	***Anchored-membrane serine/threonine-protein kinase PknF***			
pknF	R.IMLADFGIAGWVDDPSGLTATNMTVGTV**pS**(0.847)**pY**(0.847)AAPEQLMGNELDGR.A	0.99999	pY184	54%
	R.IMLADFGIAGWVDDPSGLTATNMTVG**pT**(0.362)VS**pY**(0.515)AAPEQLMGNELDGR.A	0.999987	pY184	
	R.IMLADFGIAGWVDDP**pS**(0.468)GLTATNMTVGTVS**pY**(0.407)AAPEQLMGNELDGR.A	0.999978	pY184	
	R.IMLADFGIAGWVDDPSGL**pT**(0.421)ATNMTVGTVS**pY**(0.358)AAPEQLMGNELDGR.A	0.999985	pY184	
	R.ADQ**pY**(0.863)ALAATAFHLLTGSPPFQHANPAVVISQHLSASPPAIGDR.V	0.999999	pY200	
	R.ADQ**pY**(0.667)ALAATAFHLLTGSPPFQHANPAVVISQHL**pS**(0.788)ASPPAIGDR.V	0.999999	pY200	
	R.ADQ**pY**(0.608)ALAATAFHLLTGSPPFQHANPAVVI**pS**(0.793)QHLSASPPAIGDR.V	0.999999	pY200	
Rv0410c	***Serine/threonine-protein kinase PknG***			
pknG	R.INSFG**pY**(0.524)LYGTPGFQAPEIVR.T	0.995222	pY305	88.3%
	R.**pY**(1.000)VDGLPEDDPVLK.T	0.990381	pY349	
	R.YVDGLPEDDPVLKTYDS**pY**(1.000)GR.L	0.915511	pY366	
Rv1266c	***Probable transmembrane serine/threonine-protein kinase PknH***			
pknH	R.VGSMFGP**pY**(1.000)HLK.R	0.999999	pY16	57.0%
	R.VG**pS**(1.000)MFGP**pY**(1.000)HLK.R	0.999999	pY16	
	M.GEV**pY**(0.998)EAEHTVK.E	0.995050	pY31	
	N.SGPMPASGP**pT**(0.637)PTPQ**pY**(0.730)YQGGGWGAPPSGGPSPWAQTPR.K	0.999966	pY375	
	N.SGPMPA**pS**(0.535)GPTPTPQ**pY**(0.473)YQGGGWGAPPSGGPSPWAQTPR.K	0.999973	pY375	
	N.SGPMPASGPTPTPQ**pY**(0.538)YQGGGWGAPPSGGPSPWAQTPR.K	0.999988	pY375	
	N.SGPMPASGP**pT**(0.505)PTPQY**pY**(0.680)QGGGWGAPPSGGPSPWAQTPR.K	0.999973	pY376	
	N.**pS**(0.579)GPMPASGPTPTPQY**pY**(0.538)QGGGWGAPPSGGPSPWAQTPR.K	0.959902	pY376	
	N.SGPMPASGPTP**pT**(0.448)PQY**pY**(0.524)QGGGWGAPPSGGPSPWAQTPR.K	0.976931	pY376	
	N.SGPMPA**pS**(0.425)GPTPTPQY**pY**(0.477)QGGGWGAPPSGGPSPWAQTPR.K	0.999973	pY376	
	N.SGPMPASGPTPTPQY**pY**(0.595)QGGGWGAPPSGGPSPWAQTPR.K	0.999947	pY376	
	R.QPPAPPVTPPGVQPAPKPS**pY**(0.470)TPPAQPGPAGQR.P	0.999760	pY336	

[a]Serine/Threonine protein kinase with gene name and product according to Tuberculist (http://tuberculist.epfl.ch/).
[b]pS, pT, pY indicates a phosphorylated serine, threonine or tyrosine residue.
[c]iProphet probability for each identified peptide.
[d]Phosphorylation site in protein sequence.
[e]Sequence coverage based on all identified peptides using ProteinProphet.
[f]PTMProphet score, probability for the PTM site localization.

Table 4-S1. Phosphotyrosine peptide identifications in recombinant STPKs. Identified peptides are listed with their associated proteins including the iProphet probability for each peptide and a probability for the PTM site localization. The sequence coverage was determined by ProteinProphet and includes all identified peptides.

Table S2. Phosphotyrosine peptide identifications in *Mtb* lysate

Rv number	Gene name[a]	Product	FC[b]	pY peptide sequence with PTMProphet score[c]	iProphet[d]	Sample[e]	Essentiality
Rv0020c	*fhaA*	Conserved protein with FHA domain	1	R.GGYPPETGGYPPQPG**pY**(1.000)[f]PRPR.H	0.999581	1,2,3,4,5	*In vitro*
				R.GG**pY**(0.728)PPETGGYPPQPGYPRPR.H	0.999798	5	
				P.G**pY**(0.966)APQGGGYAEPAGR.D	0.977039	4	
				R.HEEGSYVPSGPPGPPEQRPA**pY**(0.500)PDQGGYDQGYQQGATTYGR.Q	0.998198	4,5	
				R.HEEGSYVPSGPPGPPEQRPAYPDQGG**pY**(0.384)DQGYQQGATTYGR.Q	0.994956	4	
				R.HPDQGDYPEQIG**pY**(1.000)PDQGGYPEQR.G	0.995576	1,2,3,4,5	
				R.HPDQGD**pY**(0.579)PEQIGYPDQGGYPEQR.G	0.999476	4,5	
				R.QDYGGGAD**pY**(0.903)TRYTESPR.V	0.867031	1	
				R.QSGG**pY**(0.700)GPSPGGGQPGYGGYGEYGR.G	0.942068	2	
				R.VPG**pY**(1.000)APQGGGYAEPAGR.D	0.999786	1,2,3,4,5	
				R.VPGYAPQGGG**pY**(0.999)AEPAGR.D	0.999998	5	
				R.YTESPRVPG**pY**(0.999)APQGGGYAEPAGR.D	0.927423	5	
				R.**pY**(0.376)TESPRVPGYAPQGGGYAEPAGR.D	0.824251	2	
Rv0312		Conserved hypothetical proline and threonine rich protein	2[g]	R.VIDQFG**pY**(0.434)GDDVDPASTAAVGQLGQLR.E	0.887772	4	*In vitro*
Rv0361		Probable conserved membrane protein	3	R.TES**pY**(0.614)PLVPHDAETETVVITTSDNDAAVTQPEAQR.E	0.999819	1,2,3,4,5	
Rv0440	*groEL2*	60 kDa chaperonin 2 GroEL2	2	R.QEIENSDSD**pY**(0.993)DREK.L	0.937818	1,5	*In vitro*
Rv0548c	*menB*	Naphthoate synthase MenB	4	R.SGYQ**pY**(0.360)ASGDTADTVDVAR.A	0.992600	4	
Rv0876c		Possible conserved transmembrane protein	3	R.M**pY**(1.000)LLVHR.A	0.994248	2	Growth on cholesterol
				R.**pY**(1.000)LPPLGEQPEPER.S	0.973662	5	
Rv0931c	*pknD*	Transmembrane serine/threonine-protein kinase D	1	R.AASDPGLTQTGTAVGT**pY**(0.444)NYMAPER.F	0.999490	3	
Rv1326c	*glgB*	1,4-alpha-glucan branching enzyme	4	R.VDAVA**pS**M_{ox}L**pY**(0.696)LD**pY**(0.750)SRPEGGW**pT**PNVHGGR.E	0.772531	2	*In vitro*
Rv1638A		Conserved hypothetical protein	3[g]	R.**pY**(1.000)DSAGVPTFESVR.E	0.991416	1,2,3,4	
				M.PDEPTPPEATTPNSESDPR**pY**(0.450)DSAGVPTFESVR.E	0.927887	5	

Rv2198c	*mmpS3*	Probable conserved membrane protein MmpS3	3	R.AYSAPESEHVTGGP**pY**(1.000)VPADLR.L	0.999983	1,2,3,4,5	*In vitro*
				R.A**pY**(0.695)SAPESEHVTGGPYVPADLR.L	0.999472	1,3,4,5	
Rv2213	*pepB*	Probable aminopeptidase PepB	4	M.TTEPG**pY**(0.917)LSPSVAVATSMPK.R	0.999860	1,2,4	
				M.TTEPG**pY**(0.732)LSPSVAVATSM$_{ox}$PK.R	0.892905	2	
Rv2226		Conserved protein	3[g]	R.**pY**(0.551)TAAATGADNVSQEAK.V	0.995072	3	
Rv2373c	*dnaJ2*	Probable chaperone protein DnaJ2	2	K.EISVA**pY**(0.994)EVL**pS**(0.961)DPDKR.R	0.964878	4	*In vitro*
Rv2386c	*mbtI*	Isochorismate synthase MbtI	5	R.**pY**(1.000)GLQQRLAPH**pT**(1.000)PLARVF**pS**(1.000) PR.T	0.935696	4	*In vitro*
Rv2395		Probable conserved integral membrane protein	3	R.RALVTGSDLP**pY**(1.000)PEGVAGAEVLK.L	0.926200	1	
Rv3610c	*ftsH*	Membrane-bound protease FtsH	3	K.LIEAAHTEAWEILTE**pY**(0.777)R.D		5	*In vitro*
Rv3808c	*glfT2*	Bifunctional UDP-galactofuranosyl transferase GlfT2	3	K.L**pY**(1.000)LEESTTNARR.A	0.782481	3	*In vitro*

[a]Gene name, product and functional category (FC) according to Tuberculist (http://tuberculist.epfl.ch/).

[b]Functional category: 1 Regulatory proteins, 2 Virulence, detoxification, adaptation, 3 Cell wall and cell processes, 4 Intermediary metabolism and respiration, 5 Lipid Metabolism.

[c]pS, pT, pY indicates a phosphorylated serine, threonine or tyrosine residue, M_{ox} indicates oxidized methionine.

[d]iProphet probability for each identified peptide.

[e]Sample number 1-5 indicates the lysate sample the peptide was identified.

[f]PTMProphet score, probability for PTM site localization.

[g]Conserved hypotheticals, prediction of functional category based on GO and InterPro, see Mazandu and Mulder *et al. [167]*.

Table 4-S2. Phosphotyrosine peptide identifications in *Mtb* lysate. Identified peptides are listed with their associated genes, protein names, functional annotation, and essentiality according to Tuberculist. Peptide identifications are based on the Sequest search algorithm and the TPP software. An iProphet probability for each peptide is listed and the phosphorylation site localization was evaluated with a probability through PTMProphet (PTM Score).

Table S3. Phosphotyrosine peptide identifications in public available MS data on *Mtb*

Rv number	Gene name[a]	Product	Functional category	pY peptide sequence[b]	iProphet[c]	Sample[d]
A: Mtb H37Rv by Kelkar et al., PeptideAtlas PAe001767, 123 LC-MS/MS runs, 1.385.970 spectra						
Rv0020c	*fhaA*	Conserved protein with FHA domain, FhaA	regulatory proteins	R.GGYPPETGGYPPQPG**pY**(1.000)PRPR.H	0.999840	CF, OGE
Rv0206c	*mmpL3*	Possible conserved transmembrane transport protein MmpL3	cell wall and cell processes	K.K**pY**(0.347)TFVSIPLK.G	0.999222	SCX
Rv0400c	*fadE7*	Acyl-CoA dehydrogenase FadE7	lipid metabolism	R.TILGGNGISLE**pY**(1.000)PVIR.H	0.951676	CF
Rv0564c	*gpdA1*	Probable glycerol-3-phosphate dehydrogenase [NAD(P)+] GpdA1 (NAD(P)H-dependent glycerol-3-phosphate dehydrogenase) (NAD(P)H-dependent dihydroxyacetone-phosphate reductase)	lipid metabolism	R.**pY**(1.000)LGNDVVLSDTLR.A	0.950437	CF
Rv1431		Conserved membrane protein	cell wall and cell processes	K.LNHHFP**pY**(1.000)VVAVMTSNNALLR.S	0.999887	CF
Rv1908c	*katG*	Catalase-peroxidase-peroxynitritase T KatG	virulence, detoxification, adaptation	R.DLENPLAAVQM$_{ox}$GLI**pY**(1.000)VNP.E	0.965557	SCX
Rv3722c		Conserved hypothetical protein	cell wall and cell	N.**pY**(1.000)GGQHGLPGLR.A	0.997713	OGE
Rv3859c	*gltB*	Probable ferredoxin-dependent glutamate synthase [NADPH] (large subunit) GltB (L-glutamate synthase) (L-glutamate synthetase) (NADH-glutamate synthase) (glutamate synthase (NADH))(NADPH-GOGAT)	intermediary metabolism and respiration	R.EVHHM$_{ox}$AALVGFGAAAINP**pY**(1.000)LVFE**pS**(1.000)IEDM$_{ox}$LDR.G	0.999652	SCX
B: Mtb H37Rv by Schubert et al., PeptideAtlas PAe003619, 24 LC-MS/MS runs, 236.475 spectra						
Rv0020c	*fhaA*	Conserved protein with FHA domain, FhaA	regulatory proteins	R.GGYPPETGGYPPQPG**pY**(1.000)PRPR.H	0.999999	OGE
Rv1037 +4[f]	*esxI*	Putative ESAT-6 like protein EsxI	cell wall and cell processes	M.TIN**pY**(0.534)QFGDVDAHGAMIR.A	0.954463	OGE
Rv1188		Probable proline dehydrogenase	intermediary metabolism and respiration	L.GTVVQA**pY**(1.000)LR.R	0.914291	OGE
Rv1405c		Putative methyltransferase	intermediary metabolism and respiration	R.EA**pY**(1.000)V**pT**(1.000)GLLGDGV.T	0.965678	OGE

Rv1479	*moxR1*	Probable transcriptional regulatory protein MoxR1	regulatory proteins	M.TSAGGFPAGAGG**pY**(0.480)QTPGG HSASPAHEAPPGGAEGLAAEVHTLER. A	0.973951	OGE
Rv1522c	*mmpL12*	Probable conserved transmembrane transport protein MmpL12	cell wall and cell processes	A.VSVLIILILV**pY**(1.000)R.N	0.917276	OGE
Rv1798	*eccA* $_5$	ESX conserved component EccA5. ESX-5 type VII secretion system protein	cell wall and cell processes	R.**pY**(1.000)M_{ox}ADAEKNTEEMLEGA.L	0.928899	OGE
C: M. bovis BCG by Schubert et al., PeptideAtlas PAe001580, 24 LC-MS/MS runs, 213.183 spectra						
Rv0020c	*fhaA*	Conserved protein with FHA domain, FhaA	regulatory proteins	R.GGYPPETGGYPPQPG**pY**(1.000)PRP R.H	0.999999	OGE
Rv1298	*rpmE*	50S ribosomal protein L31 RpmE	information pathways	R.IVVEVCSQCHPF**pY**(0.534)TGK.Q	0.997876	OGE
Rv1479	*moxR1*	Probable transcriptional regulatory protein MoxR1	regulatory proteins	M.TSAGGFPAGAGG**pY**(0.262)QTPGG HSASPAHEAPPGGAEGLAAEVHTLER. A	0.985976	OGE
Rv3758c	*proV*	Possible osmoprotectant (glycine betaine/carnitine/choline/L-proline) transport ATP-binding protein ABC transporter ProV	virulence, detoxification, adaptation	R.LGIG**pY**(1.000)VIQNAG.L	0.999977	OGE
D: Mtb A12998 and A7494 by Huang and coworkers, Peptidome PSE133, 250 LC-MS/MS runs, 1.784.650 spectra						
Rv0006	*gyrA*	DNA gyrase (subunit A) GyrA (DNA topoisomerase (ATP-hydrolysing)) (DNA topoisomerase II) (type II DNA topoisomerase)	information pathways	R.MAQPW**pS**(1.000)LR**pY**(1.000)PLVD GQGNFGSPGNDPPAAMR.Y	0.999640	MTA12998 MudPit_FASP
Rv0020c	*fhaA*	Conserved protein with FHA domain, FhaA	regulatory proteins	R.GGYPPETGGYPPQPG**pY**(1.000)PRP R.H	0.998642	MTA7494 MudPit_SPE
Rv0020c	*fhaA*	Conserved protein with FHA domain, FhaA	regulatory proteins	R.VPG**pY**(1.000)APQGGGYAEPAGR.D	0.999999	MTA7494 MudPit_SPE
Rv0050	*ponA1*	Probable bifunctional penicillin-binding protein 1A/1B PonA1 (murein polymerase) (PBP1): penicillin-insensitive transglycosylase (peptidoglycan TGASE) + penicillin-sensitive transpeptidase (DD-transpeptidase)	cell wall and cell processes	L.**pT**(1.000)QQ**pY**(1.000)VKNALVGSAQ HGWSGLMR.K	0.931799	MTA12998 MudPit_SPE
Rv0214	*fadD4*	Probable fatty-acid-CoA ligase FadD4 (fatty-acid-CoA synthetase) (fatty-acid-CoA synthase)	lipid metabolism	R.NS**pY**(0.602)DMSSLR.R	0.987054	MTA12998 MudPit_SPE_ FASP
Rv0281		Possible S-adenosylmethionine-dependent methyltransferase	lipid metabolism	R.SHVAVEDGAPMGPDE**pY**(1.000)AAK .V	0.998283	MTA12998 MudPit_FASP

Rv0440	*groEL2*	60 kDa chaperonin 2 GroEL2 (protein CPN60-2) (GroEL protein 2) (65 kDa antigen) (heat shock protein 65) (cell wall protein A) (antigen A)	virulence, detoxification, adaptation	N.LPAGHGLNAQ**pT**(1.000)GV**pY**(1.000)EDLLAAGVADPVKVTR.S	0.995111	MTA7494 MudPit_SPE_FASP
Rv0489	*gpm1*	Probable phosphoglycerate mutase 1 Gpm1 (phosphoglyceromutase) (PGAM) (BPG-dependent PGAM)	intermediary metabolism and respiration	G.G**pT**(1.000)**pY**(1.000)LDPEAAAAGAAAVAGQGRG.-	0.999999	MTA7494 MudPit_SPE
Rv1309	*atpG*	Probable ATP synthase gamma chain AtpG	intermediary metabolism and respiration	R.V**pY**(0.999)AALLESAA**pS**(0.965)ELASR.Q	0.994021	MTA7494 MudPit_SPE
Rv1392	*metK*	Probable S-adenosylmethionine synthetase MetK (mat) (AdoMet synthetase) (methionine adenosyltransferase)	intermediary metabolism and respiration	R.VEVQVA**pY**(1.000)AIGKAAPVGLFVETFGTETEDPVKIEK.A	0.999957	MTA7494 MudPit_SPE
Rv1654	*argB*	Probable acetylglutamate kinase ArgB	intermediary metabolism and respiration	K.VVVVK**pY**(1.000)GGNAM**pT**(1.000)DD**pT**(1.000)LR.R	0.986175	MTA7494 MudPit_SPE
Rv2145c	*wag31*	Diviva family protein Wag31	cell wall and cell processes	A.F**pS**(1.000)KPPIGKRG**pY**(1.000)NEDEVDAFLDLVENELTR.L	0.999996	MTA7494 MudPit_SPE_FASP
Rv2220	*glnA1*	Glutamine synthetase GlnA1 (glutamine synthase) (GS-I)	intermediary metabolism and respiration	T.VN**pS**(1.000)**pY**(1.000)KRLVPGYEAPINLVYSQR.N	0.922671	MTA7494 MudPit_FASP
Rv2305		Unknown protein	lipid metabolism[e]	R.LLL**pS**(1.000)LARA**pY**(1.000)AGEPDEVGGPPIEEAR.N	0.999665	MTA7494_MudPit_SPE
Rv2447c	*folC*	Probable folylpolyglutamate synthase protein FolC (folylpoly-gamma-glutamate synthetase) (FPGS)	intermediary metabolism and respiration	V.D**pY**(1.000)LGADIAGIAGEK.A	0.981522	MTA12998 MudPit_SPE_FASP
Rv2524c	*fas*	Probable fatty acid synthase Fas (fatty acid synthetase)	lipid metabolism	N.PEPGR**pY**(1.000)GGEVAVVTGASK.G	0.997647	MTA12998 MudPit_SPE
Rv3151	*nuoG*	Probable NADH dehydrogenase I (chain G) NuoG (NADH-ubiquinone oxidoreductase chain G)	intermediary metabolism and respiration	R.LQDGEP**pY**(1.000)LAGTAR.T	0.979675	MTA7494 MudPit_SPE_FASP
Rv3200c		Possible transmembrane cation transporter	cell wall and cell processes	R.RVVVAVVALLL**pT**(1.000)AGIV**pY**(1.000)VDR.D	0.997957	MTA12998 MudPit_SPE
Rv3583c		Possible transcription factor	regulatory proteins	E.NAE**pY**(1.000)VGVRDVVGQEGLDKVFQVLR.A	0.994422	MTA7494 MudPit_SPE
Rv3809c	*glf*	UDP-galactopyranose mutase Glf (UDP-GALP mutase) (NAD+-flavin adenine dinucleotide-requiring enzyme)	cell wall and cell processes	K.**pY**(1.000)F**pT**(1.000)PEQARQLIAEQAAEIDTADAQNLEEK.A	0.999358	MTA12998 MudPit_SPE

[a]Gene name, product and functional category according to Tuberculist (http://tuberculist.epfl.ch/).

[b]pS, pT, pY indicates a phosphorylated serine, threonine or tyrosine residue, M_{ox} indicates oxidized methionine.

[c]iProphet probability for each identified peptide.

[d]Sample preparation, CF culture filtrate, OGE OffGel electrophoresis, SCX strong cation exchange, MudPit multidimensional protein identification technology.

[e]Conserved hypotheticals: Prediction based on GO and InterPro by Mazandu *et al.*

[f]Peptide maps to four other proteins of this family Rv1198, Rv1793, Rv2346c, Rv3619c.

Table 4-S3. *Mtb* Phosphotyrosine peptide identifications in publicly available MS data sets**.** Identified peptides are listed with their associated genes, protein names and functional annotation according to Tuberculist. An iProphet probability is listed for each peptide and the phosphorylation site localization was evaluated with a probability through PTMProphet (PTM Score).

Table S4. Tyrosine phosphorylation in *Mycobacterium smegmatis*

MSMEG ID	Product	pY peptide sequence	iProphet probability	*Mtb* orthologue
MSMEG_2958	conserved hypothetical protein	R.**pY**(0.862**)**DDRVDAPADTHVVDAPTSQIPAQPVAATATPTTSMPASQHPESMR.A	0.979839	
MSMEG_4217	DivIVA protein	R.SGAGASSQATSSIPL**pY**(0.638)EPEPEPAPAPPQPVYEAPAQPAAPQSEDTAVR.A	0.954159	Rv2145c
MSMEG_5785	probable exported protein	R.**pY**(1.000)MGIDDLLVEAPSR.D	0.999309	Rv2873
MSMEG_5196	fasciclin domain protein	V.TA**pY**(0.535)AEPETTTEAEPTVEIPDPQGPGCDDFKK.A	0.817315	Rv0817c

Table 4-S4. Tyrosine phosphorylation in *Mycobacterium smegmatis*. Unique phosphotyrosine peptides are listed with their protein names and products according to SmegmaList (http://mycobrowser.epfl.ch/smegmalist.html). pY indicates a phosphorylated tyrosine. Peptide identifications are based on a stringent, multiple search engine data analysis strategy. An iProphet probability threshold of 0.8 was applied and the phosphorylation site localization was evaluated through PTMProphet (PTM Score).

Chapter 5 – Conclusions and Future Directions

This dissertation addresses several new aspects of *Mycobacterium tuberculosis (Mtb)* phosphosignaling and protein functional annotation. In chapter 2, I detailed efforts to experimentally annotate the *Mtb* ATP binding proteins. In chapter 3, I describe the role of Ser/Thr protein kinases (STPKs) in mediating reversible oxygen-dependent replication potentially underlying latent infection. Chapter 4 describes the identification of extensive Tyr phosphorylation, dual specificity kinases, as well as initial functions of Tyr phosphorylation in *Mtb*. These studies broadly advance our understanding of the *Mtb* proteome on a functional level, identify a new molecular mechanism for the regulation of latency-associated cell adaptations, and redefine phosphosignaling in *Mtb* by discovering a previously overlooked arm of phosphorylation; however, all advances in science invariably raise new questions.

ATP-binding proteins are distributed across all functional classes, showing their importance to many cellular processes that underlie *Mtb* viability and virulence. For the purpose of functional annotation, the class of hypothetical proteins that we now identify as ATP binders is of particular interest. 72 newly identified ATP-binders are proteins unique to Mycobacteria and essential to *in vitro* growth or infection, identifying divergent ATP-dependent functional proteins that could serve as therapeutic targets for the treatment of tuberculosis (TB).

These newly annotated ATP-binders are particularly interesting because conserved elements that define ATP-binding proteins such as Walker A and B motifs are absent, presenting us with a class of non-canonical ATP-binders that can reveal new mechanisms of ATP-binding. In addition, further investigation of newly annotated ATP-binding proteins can identify previously unknown enzymatic proteins and protein families in *Mtb*. For example, Rv2179 was a hypothetical protein with no detectable sequence homology to functionally characterized proteins that labeled with the ATP activity-based probe (ATP-ABP). The Rv2179c

crystal structure revealed homology to the *E.coli* DEDD family RNase T, and follow up by *in vitro* RNase assays showed Rv2179c is in fact a functional 3' exoribonuclease [168]. Furthermore, Rv2179c orthologs are widely found in the sequenced genomes of actinobacteria and phylogenetically distant organisms such as proteobacteria, establishing a highly divergent, previously unidentified RNase family [168]. The majority of ATP-ABP labeled proteins share less than 20% sequence homology among themselves, suggesting the presence of numerous ATP-binding protein families we have yet to identify, and further characterization will reveal unknown, functional enzymes in *Mtb*.

We did not expect coverage of all ATP-binding proteins in *Mtb*. ATP-ABP only labels proteins active under the given experimental condition, aerated log phase growth, and activity of many ATP-binding proteins is likely conditional. Expanding the range of experimental conditions to include different *Mtb* growth stages, and growth under various stresses the bacilli encounters during infection such as drug treatment, low pH, starvation, and hypoxia, will provide a more comprehensive annotation of functional ATP-binding proteins. In addition, ABPP allows for direct comparative analysis of ATP-binding protein activity in different samples. For example, a recent study characterizing the differential activity of ATP-binders in aerated and hypoxic culture found 61 ATP-binding proteins with altered activity, including 38 proteins with increased activity in hypoxia [169]. Comparative analysis of condition-specific *Mtb* proteomes will characterize the dynamic range of the functional *Mtb* ATPome and potentially identify proteins that are functional solely under a given condition. Conditionally functional ATP-binding proteins are particularly interesting because they may contain conditionally-essential ATP-binders that can be used to target persistent *Mtb* populations and improve upon TB treatment.

ATP-binding proteins are one class of proteins, and comprehensive experimental annotation that includes additional enzyme classes would greatly complement our work characterizing the *Mtb* ATPome. Activity-based probes (ABPs) have been developed for a number of enzyme classes, including proteases, phosphatases, glycosidases, and

oxidoreductases. Probing the *Mtb* proteome with a variety of ABPs would provide experimental validation of predicted protein functions and help characterize protein-coding genes presently annotated as hypothetical. Ultimately, the comprehensive annotation of *Mtb* protein-coding genes and functional profiling of the *Mtb* proteome will provide the very basis for understanding *Mtb* physiology and pathogenesis.

Since ABPP relies solely on enzyme activity, one of its greatest strengths is its ability to probe and identify even the most divergent enzymes that go undetected by in silico sequence-based approaches. For example, ATP-ABP labeling of the *Mtb* proteome revealed significant divergence among ATP-binding proteins and identified a subset that lack all conserved ATP binding motifs but are still functional enzymes. The most divergent functional enzymes are very intriguing as drug targets because they represent proteins that may be unique and essential to *Mtb*, allowing for a highly targeted, specific approach to treat TB effectively with a reduced risk of toxicity. And in the case of ATP-binding proteins, there are fewer challenges to develop selective inhibitors since drug-like ATP analogs are available currently.

In addition to functional annotation, we can use ABPP to dissect kinase networks. With 11 STPKs and potential redundancy of their functions, genetic approaches to access the role of Ser/Thr phosphosignaling in growth-limiting adaptations are limited. In contrast, using the general kinase inhibitor staurosporine in combination with ABPP allows for broad screening and specific identification of STPKs that have a causal role in oxygen-dependent replication. Kinase inhibition with staurosporine resulted in decreased *Mtb* viability in reaeration and ABPP identified three STPK staurosporine targets (PknB, PknD, and PknF), providing a tractable number of STPKs to assess individual contributions to oxygen-dependent replication using genetic-based methods.

We identified PknB as a major regulator of the oxygen-dependent replication switch. Transcript and proteomic analysis revealed that PknB levels are reduced in hypoxia and restored in reaeration. Overexpression of PknB during hypoxia resulted in killing of *Mtb*, while

inhibition of PknB compromised survival in reaeration. Based on these data, we proposed a model where PknB activity drives cell growth and division in aeration but PknB activity is tightly regulated in hypoxia by protein degradation to promote bacteriostasis.

PknB specifically regulates growth in response to oxygen, as other growth-limiting stresses did not affect the survival of the *pknB* overexpressing mutant. How PknB senses changing oxygen tension is an outstanding question. Both the *Mtb* PknB kinase domain and extracellular sensor domain are essential *in vitro* and *in vivo* [170]. PknB's extracellular sensor is composed of four penicillin and Ser/Thr associated kinase domains (PASTA), suggesting that PknB senses oxygen tension indirectly [43]. In *B. subtilis*, the extracellular PASTA domains of the kinase PrkC bind peptidoglycan fragments and free muropeptides to regulate the bacilli's exit from dormancy into germination [42]. PknB also binds muropeptides and binding aids in orienting PknB to the cell poles and septum [121]. It is possible that peptidoglycan levels or structure correlate with oxygen tension, and sensing through the PknB PASTA domains places the kinase in the correct location and stimulates kinase activity to drive cell growth and division.

The functional consequences of Ser/Thr phosphorylation are ultimately determined by STPK substrates, and our work provides evidence that downstream effectors mediating oxygen-dependent replication are at least in part phosphoproteins that are PknB-specific substrates. A recent phosphoproteomic study detected 500 phosphorylation sites among 301 proteins, suggesting that each STPK has a range of substrates [48]. Importantly, many phosphorylation events were specific to different stimuli or growth conditions, providing further evidence phosphoproteins dictate bacterial adaptations [48]. Imaging of bacilli overexpressing *pknB* revealed highly elongated bacilli, suggesting downstream effectors mediating oxygen-dependent replication may be enriched for elongation and cell division proteins that modulate these processes for optimal growth. Efforts to map the PknB signaling pathway mediating oxygen-dependent replication will provide a better understanding of how *Mtb* translates hypoxia

cues into cellular responses, and may identify putative targets to inhibit *Mtb* growth for the treatment of TB.

The hypoxia and rearation model recapitulate aspects of latency and reactivation *in vitro*. In hypoxia, we found that increased PknB activity was toxic to *Mtb* and compromised the bacilli's viability, suggesting that PknB perturbation during infection may make persistent non-replicating *Mtb* more susceptible to killing. Further characterization of PknB activity affects *Mtb* virulence and viability during infection is needed. The most well established animal model for TB infection is the mouse model; however, the most commonly used mouse strains do not form necrotic, hypoxic granulomas. Recently, the C3HeB/FeJ mouse strain was reported to develop highly organized encapsulated necrotic lesions in the context of *Mtb* infection [171]. In addition, TB lesions resulting from *Mtb* infection were found to be hypoxic using a combination of positron emission tomography in live infected animals, postmortem pimonidazole immunohistochemistry, and bacterial gene expression analyses [23], making the C3HeB/FeJ mice an attractive *in vivo* model to determine how PknB activity correlates with virulence and *Mtb* viability during infection. Data from *in vivo* investigations will provide a better understanding of how *Mtb* persists under host-induced stress and of PknB as a target for non-replicating *Mtb*.

STPK-mediated reversible phosphorylation drives cellular function and adaptations in *Mtb*. We found that PknB is a critical regulator of oxygen-dependent replication, but PknB activity is dispensable for growth-limiting adaptations induced by other stresses. A recent study demonstrated that the intracellular kinase domain of PknB has unique functions in *Mtb*, and cannot be substituted by other kinase domains [170]. Furthermore, PknD is not required for oxygen-dependent regulation but mediates an osmosensory pathway that enables adaptation to osmotic stress through cell wall remodeling and virulence factor production [172]. In light of these findings, it is likely that the STPKs have stress-specific roles to overcome the dynamic microenvironments presented by the host. Hypoxia and osmotic stress represent two infection-related stresses, but *Mtb* also encounters detrimental conditions such low pH, nutrient

starvation, and nitric oxide [6,8]. Further characterization of STPK-mediated stress responses will aid in characterizing the Ser/Thr phosphosignaling network and reveal how the STPKs work in parallel, work together, and/or overlap to regulate *Mtb* adaptations that promote persistence in the host.

A major finding of this dissertation is that the *Mtb* phosphosignaling network is more extensive and complex than previously described. At least eleven two-component systems and eleven STPKs mediate phosphorylation on aspartate, histidine, serine, and threonine. Bacterial species also utilize tyrosine (Tyr) phosphorylation to drive cellular function; however, evidence that *Mtb* supports Tyr phosphorylation was inconclusive and phosphorylation on Tyr was thus thought to be absent. However, we found extensive Tyr phosphorylation in *Mtb* and showed that the STPKs function as dual specificity kinases, phosphorylating Tyr in *cis* and *trans*. In addition, Tyr phosphorylation on the activation segment regulates STPK activity *in vitro* and *in vivo*, along with Ser/Thr phosphorylation. These results reveal a significant functional role for Tyr phosphorylation in *Mtb*, and substantially expand the *Mtb* phosphosignaling network.

Ser/Thr phosphorylation on the activation segment regulates kinase activity. Sequence alignment of the eleven STPK activation segments revealed many contained a conserved C-terminal Tyr residue also conserved in the human dual specificity kinase, ERK1, and several other Tyr residues were located close to a phospho-Thr site that controls *Mtb* STPK activation. *In vitro* and *in vivo* assays showed PknB activity is regulated by phosphorylation of the conserved Tyr in the activation segment, and interestingly, PknD and PknF are phosphorylated on the same conserved Tyr. In addition, two other STPKs, PknE and PknG, showed Tyr phosphorylation on the activation segment, suggesting that many if not all STPKs are regulated by Ser, Thr, and Tyr phosphorylation. Further characterization is needed to determine how many STPKs are Tyr phosphorylated on the activation segment, and whether this Tyr phosphorylation is functional.

Dual specificity kinases can function in two distinct ways. Some dual specificity kinases autophosphorylate on Tyr but *trans* phosphorylate on Ser/Thr only, while other dual specificity kinases phosphorylate on Tyr in *cis* and *trans*. In addition to Tyr on the activation segment, we found additional Tyr phosphosites for PknB, PknD. PknE, and PknH that map to positions that are out of reach of the ATP γ-phosphate, precluding autophosphorylation and providing evidence that some STPKs are capable of *trans* phosphorylation. Mapping phospho-Tyr sites for the remaining STPKs can inform how many, if any, STPKs solely autophosphorylate on Tyr and which STPKs phosphorylate on Tyr in *cis* and *trans*. Major questions about the STPKs' ability to phosphorylate downstream substrates on Tyr and about the function of these phosphorylation events remain.

For the first time, we observed extensive *Mtb* tyrosine phosphorylation *in vivo*, identifying 63 unique sites among 49 proteins across many functional categories, indicating that Tyr phosphorylation may drive various cellular functions. In other bacterial species, Tyr phosphorylation controls growth [173-175], stress adaptation, and the best understood role is the regulation of exopolysacchride and capsule production [130-133]. Thus far, we show Tyr phosphorylation controls the activity of an essential STPK PknB known to mediate oxygen-dependent replication in *Mtb*. Moreover, the Tyr phosphorylated proteins detected by MS are enriched 2-fold for *in vitro* essentiality proteins and there are a number of cell wall biosynthesis proteins, including FhaA [110], GlfT2 [154], and GlgB [155], and the essential transporter MmpL3 that transports mycolic acids to the *Mtb* cell wall [176]. The *Mtb* Tyr phosphoproteome suggests that Tyr phosphorylation plays a similar role in *Mtb* as it does in other bacteria; however, not all phosphorylation events are functional. We provide evidence that Tyr phosphorylation mediates STPK activity, but further functional characterization is needed to determine how Tyr phosphorylation affects *Mtb* protein activity, how Tyr phosphosignaling shapes cellular function, and how this new layer of phosphorylation fits within the complex *Mtb* phosphosignaling network. As some of the *Mtb* Tyr phosphorylated proteins are being pursued

as drug targets [176-178], defining the functional role of Tyr phosphorylation can influence therapeutic design for the treatment of TB.

Reversible phosphorylation is mediated by both kinases and their cognate phosphatases. Tight regulation of phosphosignaling is critical to *Mtb* viability, as demonstrated by *Mtb's* sensitivity to slight fluctuations in PknB activity detailed in chapter 3. In *Mtb*, the sole Ser/Thr phosphatase PtsP was thought to regulate the 11 STPKs. In light of our recent evidence that STPKs function as dual specificity kinases, Tyr phosphatases must regulate STPKs in addition to PstP. *Mtb* has two functional Tyr phosphatases, PtpA and PtpB. Preliminary *in vitro* experiments indicate that PtpB may be a lipid phosphatase. And evidence clearly implicates PtpA in dephosphorylating the host protein VPs33B to prevent phagosome-lysosome fusion and *Mtb* killing during infection [7,136]. Now, it is imperative to define the intrinsic role of PtpA and PtpB in dephosphorylating of *Mtb* proteins given our findings that *Mtb* supports extensive Tyr phosphorylation, and in the case of PknB, Tyr phosphorylation is functional and serves to regulate essential protein activity. It is also important to assess the ability of PstP to regulate Tyr phosphorylation in addition to Ser/Thr phosphorylation. *Bacillus anthracis* encodes two dual specificity kinases, and the sole Ser/Thr phosphatase was subsequently shown to have dual specificity and dephosphorylate Tyr, in addition to Ser/Thr residues [55]. With incomplete annotation of the Mtb genome, it is possible that the Tyr phosphatase(s) have yet to be identified, and further investigation of non-canonical Tyr phosphatases may be needed.

Mtb and humans have co-evolved for centuries, and today TB is still one of the leading causes of death from infection worldwide. With treatment times spanning 6-9 months and drug resistance on the rise, there is a need for new, more effective treatments. The studies detailed in this dissertation advance our understanding of ATP-binding and phosphosignaling proteins underlying *Mtb* physiology and persistence, and inform future areas of investigation that may aid in next generation of TB therapeutics.

References

1. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, et al. (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A 99: 3684-3689.
2. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, et al. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393: 537-544.
3. Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 48: 77-84.
4. Sassetti CM, Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A 100: 12989-12994.
5. Brosch R, Gordon SV, Pym A, Eiglmeier K, Garnier T, et al. (2000) Comparative genomics of the mycobacteria. Int J Med Microbiol 290: 143-152.
6. Russell DG, Barry CE, 3rd, Flynn JL (2010) Tuberculosis: what we don't know can, and does, hurt us. Science 328: 852-856.
7. Bach H, Papavinasasundaram KG, Wong D, Hmama Z, Av-Gay Y (2008) Mycobacterium tuberculosis virulence is mediated by PtpA dephosphorylation of human vacuolar protein sorting 33B. Cell Host Microbe 3: 316-322.
8. Ramakrishnan L (2012) Revisiting the role of the granuloma in tuberculosis. Nat Rev Immunol 12: 352-366.
9. Kaplan G, Post FA, Moreira AL, Wainwright H, Kreiswirth BN, et al. (2003) Mycobacterium tuberculosis growth at the cavity surface: a microenvironment with failed immunity. Infect Immun 71: 7099-7108.
10. Chao MC, Rubin EJ (2010) Letting sleeping dos lie: does dormancy play a role in tuberculosis? Annu Rev Microbiol 64: 293-311.
11. Parrish NM, Dick JD, Bishai WR (1998) Mechanisms of latency in Mycobacterium tuberculosis. Trends Microbiol 6: 107-112.
12. http://www.who.int/tb/publications/global_report/en/index.html.
13. Gedde-Dahl T (1952) Tuberculous infection in the light of tuberculin matriculation. Am J Hyg 56: 139-214.
14. Marks GB, Bai J, Simpson SE, Sullivan EA, Stewart GJ (2000) Incidence of tuberculosis among a cohort of tuberculin-positive refugees in Australia: reappraising the estimates of risk. Am J Respir Crit Care Med 162: 1851-1854.
15. Comstock GW (1982) Epidemiology of tuberculosis. Am Rev Respir Dis 125: 8-15.
16. Moss AR, Hahn JA, Tulsky JP, Daley CL, Small PM, et al. (2000) Tuberculosis in the homeless. A prospective study. Am J Respir Crit Care Med 162: 460-464.
17. Selwyn PA, Hartel D, Lewis VA, Schoenbaum EE, Vermund SH, et al. (1989) A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. N Engl J Med 320: 545-550.
18. WHO (2014) Treatment of Tuberculosis: guidelines for national programmes. 4th ed.
19. Barreto ML, Pereira SM, Ferreira AA (2006) BCG vaccine: efficacy and indications for vaccination and revaccination. J Pediatr (Rio J) 82: S45-54.
20. WHO (2014) Safety of BCG vaccinei n HIV-infected chcildren.
21. Adler JJ, Rose DN (1996) Transmission and pathogenesis of tuberculosis. In: Rom WN, Garay SM, editors. Tuberculosis. Boston: Little, Brown and Co. pp. 129-140.
22. Via LE, Lin PL, Ray SM, Carrillo J, Allen SS, et al. (2008) Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. Infect Immun 76: 2333-2340.

23. Harper J, Skerry C, Davis SL, Tasneen R, Weir M, et al. (2012) Mouse model of necrotic tuberculosis granulomas develops hypoxic lesions. J Infect Dis 205: 595-602.
24. Corper HJ, Cohn ML (1951) The viability and virulence of old cultures of tubercle bacilli; studies on 30-year-old broth cultures maintained at 37 degrees C. Tubercle 32: 232-237.
25. Wayne LG (1977) Synchronized replication of Mycobacterium tuberculosis. Infect Immun 17: 528-530.
26. Wayne LG, Lin KY (1982) Glyoxylate metabolism and adaptation of Mycobacterium tuberculosis to survival under anaerobic conditions. Infect Immun 37: 1042-1049.
27. Wayne LG, Hayes LG (1996) An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. Infect Immun 64: 2062-2069.
28. Rustad TR, Harrell MI, Liao R, Sherman DR (2008) The enduring hypoxic response of Mycobacterium tuberculosis. PLoS One 3: e1502.
29. Sherrid AM, Rustad TR, Cangelosi GA, Sherman DR (2010) Characterization of a Clp protease gene regulator and the reaeration response in Mycobacterium tuberculosis. PLoS One 5: e11622.
30. Rodrigue S, Provvedi R, Jacques PE, Gaudreau L, Manganelli R (2006) The sigma factors of Mycobacterium tuberculosis. FEMS Microbiol Rev 30: 926-941.
31. Lew JM, Kapopoulou A, Jones LM, Cole ST (2011) TubercuList--10 years after. Tuberculosis (Edinb) 91: 1-7.
32. Cravatt BF, Wright AT, Kozarich JW (2008) Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. Annu Rev Biochem 77: 383-414.
33. Barglow KT, Cravatt BF (2007) Activity-based protein profiling for the functional annotation of enzymes. Nat Methods 4: 822-827.
34. Magnet S, Hartkoorn RC, Szekely R, Pato J, Triccas JA, et al. (2010) Leads for antitubercular compounds from kinase inhibitor library screens. Tuberculosis (Edinb) 90: 354-360.
35. Schreiber M, Res I, Matter A (2009) Protein kinases as antibacterial targets. Curr Opin Cell Biol 21: 325-330.
36. Adams KN, Szumowski JD, Ramakrishnan L (2014) Verapamil, and Its Metabolite Norverapamil, Inhibit Macrophage-induced, Bacterial Efflux Pump-mediated Tolerance to Multiple Anti-tubercular Drugs. J Infect Dis.
37. Adams KN, Takaki K, Connolly LE, Wiedenhoft H, Winglee K, et al. (2011) Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. Cell 145: 39-53.
38. Alber T (2009) Signaling mechanisms of the Mycobacterium tuberculosis receptor Ser/Thr protein kinases. Curr Opin Struct Biol 19: 650-657.
39. Perez J, Castaneda-Garcia A, Jenke-Kodama H, Muller R, Munoz-Dorado J (2008) Eukaryotic-like protein kinases in the prokaryotes and the myxobacterial kinome. Proc Natl Acad Sci U S A 105: 15950-15955.
40. Av-Gay Y, Everett M (2000) The eukaryotic-like Ser/Thr protein kinases of Mycobacterium tuberculosis. Trends Microbiol 8: 238-244.
41. Greenstein AE, Grundner C, Echols N, Gay LM, Lombana TN, et al. (2005) Structure/function studies of Ser/Thr and Tyr protein phosphorylation in Mycobacterium tuberculosis. J Mol Microbiol Biotechnol 9: 167-181.
42. Shah IM, Laaberki MH, Popham DL, Dworkin J (2008) A eukaryotic-like Ser/Thr kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. Cell 135: 486-496.

43. Barthe P, Mukamolova GV, Roumestand C, Cohen-Gonsaud M (2010) The structure of PknB extracellular PASTA domain from mycobacterium tuberculosis suggests a ligand-dependent kinase activation. Structure 18: 606-615.
44. Kang CM, Abbott DW, Park ST, Dascher CC, Cantley LC, et al. (2005) The Mycobacterium tuberculosis serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape. Genes Dev 19: 1692-1704.
45. Kang CM, Nyayapathy S, Lee JY, Suh JW, Husson RN (2008) Wag31, a homologue of the cell division protein DivIVA, regulates growth, morphology and polar cell wall synthesis in mycobacteria. Microbiology 154: 725-735.
46. Park ST, Kang CM, Husson RN (2008) Regulation of the SigH stress response regulon by an essential protein kinase in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 105: 13105-13110.
47. Gopalaswamy R, Narayanan S, Chen B, Jacobs WR, Av-Gay Y (2009) The serine/threonine protein kinase PknI controls the growth of Mycobacterium tuberculosis upon infection. FEMS Microbiol Lett 295: 23-29.
48. Prisic S, Dankwa S, Schwartz D, Chou MF, Locasale JW, et al. (2010) Extensive phosphorylation with overlapping specificity by Mycobacterium tuberculosis serine/threonine protein kinases. Proc Natl Acad Sci U S A 107: 7521-7526.
49. Grangeasse C, Nessler S, Mijakovic I (2012) Bacterial tyrosine kinases: evolution, biological function and structural insights. Philos Trans R Soc Lond B Biol Sci 367: 2640-2655.
50. Whitmore SE, Lamont RJ (2012) Tyrosine phosphorylation and bacterial virulence. Int J Oral Sci 4: 1-6.
51. Hansen AM, Chaerkady R, Sharma J, Diaz-Mejia JJ, Tyagi N, et al. (2013) The Escherichia coli phosphotyrosine proteome relates to core pathways and virulence. PLoS Pathog 9: e1003403.
52. Aranda S, Laguna A, de la Luna S (2011) DYRK family of protein kinases: evolutionary relationships, biochemical properties, and functional roles. FASEB J 25: 449-462.
53. Becker W, Joost HG (1999) Structural and functional characteristics of Dyrk, a novel subfamily of protein kinases with dual specificity. Prog Nucleic Acid Res Mol Biol 62: 1-17.
54. Dhanasekaran N, Premkumar Reddy E (1998) Signaling by dual specificity kinases. Oncogene 17: 1447-1455.
55. Arora G, Sajid A, Arulanandh MD, Singhal A, Mattoo AR, et al. (2012) Unveiling the novel dual specificity protein kinases in Bacillus anthracis: identification of the first prokaryotic dual specificity tyrosine phosphorylation-regulated kinase (DYRK)-like kinase. J Biol Chem 287: 26749-26763.
56. Johnson DL, Mahony JB (2007) Chlamydophila pneumoniae PknD exhibits dual amino acid specificity and phosphorylates Cpn0712, a putative type III secretion YscD homolog. J Bacteriol 189: 7549-7555.
57. Koul A, Choidas A, Treder M, Tyagi AK, Drlica K, et al. (2000) Cloning and characterization of secretory tyrosine phosphatases of Mycobacterium tuberculosis. J Bacteriol 182: 5425-5432.
58. Bach H, Wong D, Av-Gay Y (2009) Mycobacterium tuberculosis PtkA is a novel protein tyrosine kinase whose substrate is PtpA. Biochem J 420: 155-160.
59. Chow K, Ng D, Stokes R, Johnson P (1994) Protein tyrosine phosphorylation in Mycobacterium tuberculosis. FEMS Microbiol Lett 124: 203-207.
60. Bork P (2000) Powers and pitfalls in sequence analysis: the 70% hurdle. Genome Res 10: 398-400.
61. Bork P, Koonin EV (1998) Predicting functions from protein sequences--where are the bottlenecks? Nat Genet 18: 313-318.

62. Huynen M, Snel B, Lathe W, 3rd, Bork P (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. Genome Res 10: 1204-1210.
63. Simon GM, Cravatt BF (2010) Activity-based proteomics of enzyme superfamilies: serine hydrolases as a case study. J Biol Chem 285: 11051-11055.
64. Patricelli MP, Szardenings AK, Liyanage M, Nomanbhoy TK, Wu M, et al. (2007) Functional interrogation of the kinome using nucleotide acyl phosphates. Biochemistry 46: 350-358.
65. Qiu H, Wang Y (2007) Probing adenosine nucleotide-binding proteins with an affinity-labeled nucleotide probe and mass spectrometry. Anal Chem 79: 5547-5556.
66. DiSabato G, and Jencks, W.P. (1961) Mechanism and catalysis of reactions of acyl phosphates J Am Chem 83: 4393-4400.
67. Kluger R (2000) Acyl phosphate esters: charge-directed acylation and artificial blood Synlett 12: 1708-1720.
68. Sadler NC, Angel TE, Lewis MP, Pederson LM, Chauvigne-Hines LM, et al. (2012) Activity-based protein profiling reveals mitochondrial oxidative enzyme impairment and restoration in diet-induced obese mice. PLoS One 7: e47996.
69. Speers AE, Adam GC, Cravatt BF (2003) Activity-based protein profiling in vivo using a copper(i)-catalyzed azide-alkyne [3 + 2] cycloaddition. J Am Chem Soc 125: 4686-4687.
70. Speers AE, Cravatt BF (2005) A tandem orthogonal proteolysis strategy for high-content chemical proteomics. J Am Chem Soc 127: 10018-10019.
71. Zimmer JS, Monroe ME, Qian WJ, Smith RD (2006) Advances in proteomics data analysis and display using an accurate mass and time tag approach. Mass Spectrom Rev 25: 450-482.
72. Gillespie JJ, Wattam AR, Cammer SA, Gabbard JL, Shukla MP, et al. (2011) PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. Infect Immun 79: 4286-4298.
73. Reddy TB, Riley R, Wymore F, Montgomery P, DeCaprio D, et al. (2009) TB database: an integrated platform for tuberculosis research. Nucleic Acids Res 37: D499-508.
74. Haft DH, Selengut JD, White O (2003) The TIGRFAMs database of protein families. Nucleic Acids Res 31: 371-373.
75. Doerks T, van Noort V, Minguez P, Bork P (2012) Annotation of the M. tuberculosis hypothetical orfeome: adding functional information to more than half of the uncharacterized proteins. PLoS One 7: e34302.
76. Hughes KT, Mathee K (1998) The anti-sigma factors. Annu Rev Microbiol 52: 231-286.
77. Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4: 363-371.
78. Greenstein AE, MacGurn JA, Baer CE, Falick AM, Cox JS, et al. (2007) M. tuberculosis Ser/Thr protein kinase D phosphorylates an anti-anti-sigma factor homolog. PLoS Pathog 3: e49.
79. Fortune SM, Jaeger A, Sarracino DA, Chase MR, Sassetti CM, et al. (2005) Mutually dependent secretion of proteins required for mycobacterial virulence. Proc Natl Acad Sci U S A 102: 10676-10681.
80. MacGurn JA, Raghavan S, Stanley SA, Cox JS (2005) A non-RD1 gene cluster is required for Snm secretion in Mycobacterium tuberculosis. Mol Microbiol 57: 1653-1663.
81. Newton GL, Leung SS, Wakabayashi JI, Rawat M, Fahey RC (2011) The DinB superfamily includes novel mycothiol, bacillithiol, and glutathione S-transferases. Biochemistry 50: 10751-10760.

82. Lombana TN, Echols N, Good MC, Thomsen ND, Ng HL, et al. (2010) Allosteric activation mechanism of the Mycobacterium tuberculosis receptor Ser/Thr protein kinase, PknB. Structure 18: 1667-1677.
83. Ansong C, Purvine SO, Adkins JN, Lipton MS, Smith RD (2008) Proteogenomics: needs and roles to be filled by proteomics in genome annotation. Brief Funct Genomic Proteomic 7: 50-62.
84. Venter E, Smith RD, Payne SH (2011) Proteogenomic analysis of bacteria and archaea: a 46 organism case study. PLoS One 6: e27587.
85. Galperin MY, Koonin EV (2010) From complete genome sequence to 'complete' understanding? Trends Biotechnol 28: 398-406.
86. Deutschbauer A, Price MN, Wetmore KM, Shao W, Baumohl JK, et al. (2011) Evidence-based annotation of gene function in Shewanella oneidensis MR-1 using genome-wide fitness profiling across 121 conditions. PLoS Genet 7: e1002385.
87. Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, et al. (2011) EcoCyc: a comprehensive database of Escherichia coli biology. Nucleic Acids Res 39: D583-590.
88. Rost B (2002) Enzyme function less conserved than anticipated. J Mol Biol 318: 595-608.
89. Price MN, Dehal PS, Arkin AP (2007) Orthologous transcription factors in bacteria have different functions and regulate different genes. PLoS Comput Biol 3: 1739-1750.
90. Ren Q, Paulsen IT (2005) Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. PLoS Comput Biol 1: e27.
91. Kelkar DS, Kumar D, Kumar P, Balakrishnan L, Muthusamy B, et al. (2011) Proteogenomic analysis of Mycobacterium tuberculosis by high resolution mass spectrometry. Mol Cell Proteomics 10: M111 011627.
92. Gomez JE, Clatworthy A, Hung DT (2011) Probing bacterial pathogenesis with genetics, genomics, and chemical biology: past, present, and future approaches. Crit Rev Biochem Mol Biol 46: 41-66.
93. Mawuenyega KG, Forst CV, Dobos KM, Belisle JT, Chen J, et al. (2005) Mycobacterium tuberculosis functional network analysis by global subcellular protein profiling. Mol Biol Cell 16: 396-404.
94. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, et al. (2006) TM4 microarray software suite. Methods Enzymol 411: 134-193.
95. Kelly RT, Page JS, Luo Q, Moore RJ, Orton DJ, et al. (2006) Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. Anal Chem 78: 7796-7801.
96. Yates JR, 3rd, Eng JK, McCormack AL, Schieltz D (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal Chem 67: 1426-1436.
97. Kim S, Gupta N, Pevzner PA (2008) Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. J Proteome Res 7: 3354-3363.
98. Jaitly N, Mayampurath A, Littlefield K, Adkins JN, Anderson GA, et al. (2009) Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. BMC Bioinformatics 10: 87.
99. Monroe ME, Tolic N, Jaitly N, Shaw JL, Adkins JN, et al. (2007) VIPER: an advanced software package to support high-throughput LC-MS peptide identification. Bioinformatics 23: 2021-2023.

100. Stanley JR, Adkins JN, Slysz GW, Monroe ME, Purvine SO, et al. (2011) A statistical method for assessing peptide identification confidence in accurate mass and time tag proteomics. Anal Chem 83: 6135-6140.
101. Polpitiya AD, Qian WJ, Jaitly N, Petyuk VA, Adkins JN, et al. (2008) DAnTE: a statistical tool for quantitative analysis of -omics data. Bioinformatics 24: 1556-1558.
102. Getahun H, Gunneberg C, Granich R, Nunn P (2010) HIV infection-associated tuberculosis: the epidemiology and the response. Clin Infect Dis 50 Suppl 3: S201-207.
103. Rustad TR, Sherrid AM, Minch KJ, Sherman DR (2009) Hypoxia: a window into Mycobacterium tuberculosis latency. Cell Microbiol 11: 1151-1159.
104. Novy FG, Soule MH (1925) Microbic Respiration. II. Repiration of the Tubercle Bacillus. J Infect Dis 36: 168.
105. Rich AR, Follis RH, Jr. (1942) The effect of low oxygen tension upon the development of experimental tuberculosis. Bull Johns Hopkins Hosp 71: 345-363.
106. Tsai MC, Chakravarty S, Zhu G, Xu J, Tanaka K, et al. (2006) Characterization of the tuberculous granuloma in murine and human lungs: cellular composition and relative tissue oxygen tension. Cell Microbiol 8: 218-232.
107. Boshoff HI, Barry CE, 3rd (2005) Tuberculosis - metabolism and respiration in the absence of growth. Nat Rev Microbiol 3: 70-80.
108. Flynn JL, Chan J (2001) Tuberculosis: latency and reactivation. Infect Immun 69: 4195-4201.
109. Gomez JE, McKinney JD (2004) *M. tuberculosis* persistence, latency, and drug tolerance. Tuberculosis (Edinb) 84: 29-44.
110. Gee CL, Papavinasasundaram KG, Blair SR, Baer CE, Falick AM, et al. (2012) A phosphorylated pseudokinase complex controls cell wall synthesis in mycobacteria. Sci Signal 5: ra7.
111. Fernandez P, Saint-Joanis B, Barilone N, Jackson M, Gicquel B, et al. (2006) The Ser/Thr protein kinase PknB is essential for sustaining mycobacterial growth. J Bacteriol 188: 7778-7784.
112. Fischer JJ, Graebner Baessler OY, Dalhoff C, Michaelis S, Schrey AK, et al. (2010) Comprehensive identification of staurosporine-binding kinases in the hepatocyte cell line HepG2 using Capture Compound Mass Spectrometry (CCMS). J Proteome Res 9: 806-817.
113. Wright AT, Cravatt BF (2007) Chemical proteomic probes for profiling cytochrome p450 activities and drug interactions in vivo. Chem Biol 14: 1043-1051.
114. Ansong C, Ortega C, Payne SH, Haft DH, Chauvigne-Hines LM, et al. (2013) Identification of widespread adenosine nucleotide binding in Mycobacterium tuberculosis. Chem Biol 20: 123-133.
115. Smith RD, Anderson GA, Lipton MS, Pasa-Tolic L, Shen Y, et al. (2002) An accurate mass tag strategy for quantitative and high-throughput proteome measurements. Proteomics 2: 513-523.
116. http://www.beiresources.org**.**
117. Chao JD, Papavinasasundaram KG, Zheng X, Chavez-Steenbock A, Wang X, et al. (2010) Convergence of Ser/Thr and two-component signaling to coordinate expression of the dormancy regulon in Mycobacterium tuberculosis. J Biol Chem 285: 29239-29246.
118. Galagan JE, Minch K, Peterson M, Lyubetskaya A, Azizi E, et al. (2013) The Mycobacterium tuberculosis regulatory network and hypoxia. Nature 499: 178-183.
119. van Kessel JC, Hatfull GF (2007) Recombineering in Mycobacterium tuberculosis. Nat Methods 4: 147-152.

120. Duran R, Villarino A, Bellinzoni M, Wehenkel A, Fernandez P, et al. (2005) Conserved autophosphorylation pattern in activation loops and juxtamembrane regions of Mycobacterium tuberculosis Ser/Thr protein kinases. Biochem Biophys Res Commun 333: 858-867.
121. Mir M, Asong J, Li X, Cardot J, Boons GJ, 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: e1002182.
122. Szekely R, Waczek F, Szabadkai I, Nemeth G, Hegymegi-Barakonyi B, et al. (2008) A novel drug discovery concept for tuberculosis: inhibition of bacterial and host cell signalling. Immunol Lett 116: 225-231.
123. Lougheed KE, Osborne SA, Saxty B, Whalley D, Chapman T, et al. (2011) Effective inhibitors of the essential kinase PknB and their potential as anti-mycobacterial agents. Tuberculosis (Edinb) 91: 277-286.
124. Chapman TM, Bouloc N, Buxton RS, Chugh J, Lougheed KE, et al. (2012) Substituted aminopyrimidine protein kinase B (PknB) inhibitors show activity against Mycobacterium tuberculosis. Bioorg Med Chem Lett 22: 3349-3353.
125. Ruan BF, Zhu HL (2012) The chemistry and biology of the bryostatins: potential PKC inhibitors in clinical development. Curr Med Chem 19: 2652-2664.
126. Wards BJ, Collins DM (1996) Electroporation at elevated temperatures substantially improves transformation efficiency of slow-growing mycobacteria. FEMS Microbiol Lett 145: 101-105.
127. Voskuil MI, Schnappinger D, Visconti KC, Harrell MI, Dolganov GM, et al. (2003) Inhibition of respiration by nitric oxide induces a Mycobacterium tuberculosis dormancy program. J Exp Med 198: 705-713.
128. Vandal OH, Nathan CF, Ehrt S (2009) Acid resistance in Mycobacterium tuberculosis. J Bacteriol 191: 4714-4721.
129. Wehenkel A, Bellinzoni M, Grana M, Duran R, Villarino A, et al. (2008) Mycobacterial Ser/Thr protein kinases and phosphatases: physiological roles and therapeutic potential. Biochim Biophys Acta 1784: 193-202.
130. Bender MH, Cartee RT, Yother J (2003) Positive correlation between tyrosine phosphorylation of CpsD and capsular polysaccharide production in Streptococcus pneumoniae. J Bacteriol 185: 6057-6066.
131. Ilan O, Bloch Y, Frankel G, Ullrich H, Geider K, et al. (1999) Protein tyrosine kinases in bacterial pathogens are associated with virulence and production of exopolysaccharide. EMBO J 18: 3241-3248.
132. Vincent C, Duclos B, Grangeasse C, Vaganay E, Riberty M, et al. (2000) Relationship between exopolysaccharide production and protein-tyrosine phosphorylation in gram-negative bacteria. J Mol Biol 304: 311-321.
133. Wugeditsch T, Paiment A, Hocking J, Drummelsmith J, Forrester C, et al. (2001) Phosphorylation of Wzc, a tyrosine autokinase, is essential for assembly of group 1 capsular polysaccharides in Escherichia coli. J Biol Chem 276: 2361-2371.
134. Gnad F, Gunawardena J, Mann M (2011) PHOSIDA 2011: the posttranslational modification database. Nucleic Acids Res 39: D253-260.
135. Lee DC, Jia Z (2009) Emerging structural insights into bacterial tyrosine kinases. Trends Biochem Sci 34: 351-357.
136. Wong D, Chao JD, Av-Gay Y (2013) Mycobacterium tuberculosis-secreted phosphatases: from pathogenesis to targets for TB drug development. Trends Microbiol 21: 100-109.
137. Ballif BA, Villen J, Beausoleil SA, Schwartz D, Gygi SP (2004) Phosphoproteomic analysis of the developing mouse brain. Mol Cell Proteomics 3: 1093-1101.

138. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, et al. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127: 635-648.
139. Ge R, Shan W (2011) Bacterial phosphoproteomic analysis reveals the correlation between protein phosphorylation and bacterial pathogenicity. Genomics Proteomics Bioinformatics 9: 119-127.
140. Macek B, Gnad F, Soufi B, Kumar C, Olsen JV, et al. (2008) Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. Mol Cell Proteomics 7: 299-307.
141. Jadeau F, Grangeasse C, Shi L, Mijakovic I, Deleage G, et al. (2012) BYKdb: the Bacterial protein tYrosine Kinase database. Nucleic Acids Res 40: D321-324.
142. Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, et al. (2010) A guided tour of the Trans-Proteomic Pipeline. PROTEOMICS 10: 1150-1159.
143. Kapp EA, Schütz F, Reid GE, Eddes JS, Moritz RL, et al. (2003) Mining a Tandem Mass Spectrometry Database To Determine the Trends and Global Factors Influencing Peptide Fragmentation. Analytical Chemistry 75: 6251-6264.
144. Olsen JV, Macek B, Lange O, Makarov A, Horning S, et al. (2007) Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods 4: 709-712.
145. Zhang G, Neubert TA (2011) Comparison of three quantitative phosphoproteomic strategies to study receptor tyrosine kinase signaling. J Proteome Res 10: 5454-5462.
146. Grundner C, Gay LM, Alber T (2005) Mycobacterium tuberculosis serine/threonine kinases PknB, PknD, PknE, and PknF phosphorylate multiple FHA domains. Protein Sci 14: 1918-1921.
147. Roumestand C, Leiba J, Galophe N, Margeat E, Padilla A, et al. (2011) Structural insight into the Mycobacterium tuberculosis Rv0020c protein and its interaction with the PknB kinase. Structure 19: 1525-1534.
148. Macek B, Mijakovic I, Olsen JV, Gnad F, Kumar C, et al. (2007) The serine/threonine/tyrosine phosphoproteome of the model bacterium Bacillus subtilis. Mol Cell Proteomics 6: 697-707.
149. Kelkar DS, Kumar D, Kumar P, Balakrishnan L, Muthusamy B, et al. (2011) Proteogenomic Analysis of Mycobacterium tuberculosis By High Resolution Mass Spectrometry. Molecular & Cellular Proteomics 10.
150. Schubert Olga T, Mouritsen J, Ludwig C, Röst Hannes L, Rosenberger G, et al. (2013) The Mtb Proteome Library: A Resource of Assays to Quantify the Complete Proteome of Mycobacterium tuberculosis. Cell Host & Microbe 13: 602-612.
151. Villarino A, Duran R, Wehenkel A, Fernandez P, England P, et al. (2005) Proteomic identification of M. tuberculosis protein kinase substrates: PknB recruits GarA, a FHA domain-containing protein, through activation loop-mediated interactions. J Mol Biol 350: 953-963.
152. Ortega C, Liao R, Anderson L, Rustad T, Ollodart A, et al. (2013) Mycobacterium tuberculosis Ser/Thr protein kinase B mediates an oxygen-induced replication switch PLoS Biol in press.
153. Grundner C, Cox JS, Alber T (2008) Protein tyrosine phosphatase PtpA is not required for Mycobacterium tuberculosis growth in mice. FEMS Microbiol Lett 287: 181-184.
154. Kremer L, Dover LG, Morehouse C, Hitchin P, Everett M, et al. (2001) Galactan biosynthesis in Mycobacterium tuberculosis. Identification of a bifunctional UDP-galactofuranosyltransferase. J Biol Chem 276: 26430-26440.
155. Sambou T, Dinadayala P, Stadthagen G, Barilone N, Bordat Y, et al. (2008) Capsular glucan and intracellular glycogen of Mycobacterium tuberculosis: biosynthesis and impact on the persistence in mice. Mol Microbiol 70: 762-774.

156. Villen J, Gygi SP (2008) The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protocols 3: 1630-1638.
157. Kessner D, Chambers M, Burke R, Agus D, Mallick P (2008) ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24: 2534-2536.
158. Martens L, Chambers M, Sturm M, Kessner D, Levander F, et al. (2011) mzML - a Community Standard for Mass Spectrometry Data. Molecular & Cellular Proteomics 10: R110.000133.
159. Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. Analytical Chemistry 74: 5383-5392.
160. Shteynberg D, Deutsch EW, Lam H, Eng JK, Sun Z, et al. (2011) iProphet: Multi-level Integrative Analysis of Shotgun Proteomic Data Improves Peptide and Protein Identification Rates and Error Estimates. Molecular & Cellular Proteomics 10: M111.007690.
161. Eng JK, Jahan TA, Hoopmann MR (2013) Comet: An open-source MS/MS sequence database search tool. PROTEOMICS 13: 22-24.
162. Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20: 1466-1467.
163. Keller A, Eng J, Zhang N, Li X-j, Aebersold R (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol Syst Biol 1: 2005.0017.
164. Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003) A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. Analytical Chemistry 75: 4646-4658.
165. Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20: 1466-1467.
166. Kelkar DS, Kumar D, Kumar P, Balakrishnan L, Muthusamy B, et al. (2011) Proteogenomic Analysis of Mycobacterium tuberculosis By High Resolution Mass Spectrometry. Molecular & Cellular Proteomics 10: M111.011627.
167. Mazandu GK, Mulder NJ (2012) Function Prediction and Analysis of Mycobacterium tuberculosis Hypothetical Proteins. International Journal of Molecular Sciences 13: 7283-7302.
168. Abendroth J, Ollodart A, Andrews ES, Myler PJ, Staker BL, et al. (2014) Mycobacterium tuberculosis Rv2179c protein establishes a new exoribonuclease family with broad phylogenetic distribution. J Biol Chem 289: 2139-2147.
169. Wolfe LM, Veeraraghavan U, Idicula-Thomas S, Schurer S, Wennerberg K, et al. (2013) A chemical proteomics approach to profiling the ATP-binding proteome of Mycobacterium tuberculosis. Mol Cell Proteomics 12: 1644-1660.
170. Chawla Y, Upadhyay SK, Khan S, Nagarajan SN, Forti F, et al. (2014) Protein Kinase B (PknB) of Mycobacterium tuberculosis is essential for growth of the pathogen in vitro as well as for survival within the host. J Biol Chem.
171. Pichugin AV, Yan BS, Sloutsky A, Kobzik L, Kramnik I (2009) Dominant role of the sst1 locus in pathogenesis of necrotizing lung granulomas during chronic tuberculosis infection and reactivation in genetically resistant hosts. Am J Pathol 174: 2190-2201.
172. Hatzios SK, Baer CE, Rustad TR, Siegrist MS, Pang JM, et al. (2013) Osmosensory signaling in Mycobacterium tuberculosis mediated by a eukaryotic-like Ser/Thr protein kinase. Proc Natl Acad Sci U S A 110: E5069-5077.
173. Wu J, Ohta N, Zhao JL, Newton A (1999) A novel bacterial tyrosine kinase essential for cell division and differentiation. Proc Natl Acad Sci U S A 96: 13068-13073.

174. Thomasson B, Link J, Stassinopoulos AG, Burke N, Plamann L, et al. (2002) MglA, a small GTPase, interacts with a tyrosine kinase to control type IV pili-mediated motility and development of Myxococcus xanthus. Mol Microbiol 46: 1399-1413.

175. Kumagai T, Kihara H, Watanabe W, Noda M, Matoba Y, et al. (2009) A novel tyrosine-phosphorylated protein inhibiting the growth of Streptomyces cells. Biochem Biophys Res Commun 385: 534-538.

176. Grzegorzewicz AE, Pham H, Gundi VA, Scherman MS, North EJ, et al. (2012) Inhibition of mycolic acid transport across the Mycobacterium tuberculosis plasma membrane. Nat Chem Biol 8: 334-341.

177. Li K, Schurig-Briccio LA, Feng X, Upadhyay A, Pujari V, et al. (2014) Multitarget Drug Discovery for Tuberculosis and Other Infectious Diseases. J Med Chem.

178. Rayasam GV (2014) MmpL3 a potential new target for development of novel anti-tuberculosis drugs. Expert Opin Ther Targets 18: 247-256.