The *Mycobacterium tuberculosis* protein *O*-phosphorylation landscape

Andrew Frando

A dissertation submitted in partial fulfillment of the

Requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading committee:

Christoph Grundner, Chair

David Sherman

Kevin Hybiske

Program authorized to offer degree:

Pathobiology

University of Washington

Abstract

The *Mycobacterium tuberculosis* protein *O*-phosphorylation landscape

Andrew Frando

Chair of the supervisory committee

Professor Christoph Grundner

Department of Pathobiology

Protein phosphorylation is a main mechanism for translating extracellular signals into cellular adaptations. In bacteria, the two-component system has been the paradigm of protein phosphorylation. Increasingly, however, protein serine/threonine and tyrosine phosphorylation (*O*-phosphorylation) mediated by serine/threonine protein kinases (STPKs) is identified. *Mycobacterium tuberculosis* (*Mtb*) in particular has a larger repertoire of both STPKs and *O*-phosphoproteins than most bacteria, suggesting a more prevalent role of STPKs in *Mtb*. Many studies have identified individual STPK functions and substrates, but a systems-level understanding and the full scope of *O*-phosphorylation in bacteria in general and *Mtb* in particular remains unknown. In this dissertation, we aimed to establish a systems-wide understanding of *Mtb O*-phosphorylation. To this end, we combined kinase loss-of-function (LoF) and gain-of-

function (GoF) mutants with mass spectrometry (MS)-based phosphoproteomics. We found that the *Mtb* phosphoproteome is 5x larger than was previously reported. By identifying hypo-phosphorylated proteins in the LoF mutants and hyper-phosphorylated protein in the GoF mutants, we comprehensively identified over 1,400 STPK-substrates in the bacterium, assigning putative STPK substrates and functions. We showed that the *O*-phosphoproteome is organized as a higher-order network of a complexity that has previously only been associated with eukaryotes. Transcription factors (TFs) were highly phosphorylated, suggesting a large functional interface between *O*-phosphorylation and transcription. We further characterized this interface by using transcriptomics and discovered large transcriptional effects of individual STPKs. We used these data to predict functional phosphorylation events and validated these predictions for the putative zinc regulator Zur. These data provide the deepest bacterial *O*-phosphoproteome to date, identify a complex *O*-phosphorylation network, and provide a resource to assign thousands of specific phosphorylation events to individual STPKs and their transcriptional effects. Together, these data challenge the paradigm of TCSs as the main bacterial phosphosignaling mechanism and suggest that *Mtb* regulates a large swath of physiology through O-phosphorylation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

TB = Tuberculosis

Mtb = *Mycobacterium tuberculosis*

MDR-TB = multi-drug resistant TB

XDR-TB = extensively drug resistant TB

TDR-TB = totally drug resistant TB

TCS = two-component system

HK = histidine kinase

RR = response regulator

MAPK = mitogen-activated protein kinase

STPK = serine/threonine protein kinase

WT = wild-type

LoF = loss-of-function

KO = knockout

OE = overexpression

GoF = gain-of-function

LC = liquid chromatography

SRM = selected reaction monitoring

IPTG = isopropyl β-D-1-thigalactopyranoside

SDS-PAGE = sodium dodecyl sulphate-polyacrylamide gel electrophoresis

EMSA = electrophoretic mobility shift assay

CRISPRi = CRISPR interference

TMT = tandem mass tag

LF = label-free

KEGG = Kyoto encyclopedia of genes and genomes

GO = gene ontology

TF = transcription factor

DEG = differentially expressed gene

Dedication

To my loves Tyler, Cielo, and Titus,

for your unconditional love and support.

Acknowledgements

"You are only as tall as the shoulders you stand on" is a phrase I heard often during my time in the UC Berkley Filipino community. The many opportunities I've been given and lessons I've learned in life, especially during this PhD, is owed in large part to the academic and emotional support I received from the following individuals.

I would like to thank my mentor Christoph Grundner for taking a chance on me and giving me the opportunity to work on an amazing and rewarding project that continually excites me. Christoph shaped me into a better scientist by challenging me to constantly think critically, use new techniques, and improve my communication skills. I am proud of the work we produced and look forward to the future of this project and the lab. Outside of research, I've always looked forward to discussing details from our trips or life in general. Also, thanks to Christoph always appreciating my many homemade cakes and treats.

From the Grundner lab, thank you to Vishant, Claude, and Neil and previous members Corrie, Anja, Maryana, and Fablina for great scientific conversations and assistance during experiments. I would also like to thank the Sherman lab, especially Jessica Farrow-Johnson, Jessica Winkler, Tige, Reiling, and Kyle, for teaching me the basics for working with *Mtb*. A special thanks to Jon Jacobs and PNNL for their contributions to my dissertation.

I received immense support from the Seattle scientific community. I would like to thank my thesis committee members David Sherman, Nina Salama, and Josh Woodward for their input and support during my PhD. I would like to extend further thanks to David Sherman and Kevin Hybiske for reading my dissertation. I am grateful to the CGIDR and Pathobiology students, staff, and faculty that have supported me throughout this process.

I would not have made it this far without guidance from my previous mentors. Thank you to Eva Harris and Matthias Marti for introducing me to infectious disease research and inspiring me to pursue research as a career. I am thankful to Joanne Engel and Cheri Elwell for encouraging me to pursue graduate school and teaching me basic technical skills, critical thinking, and communication that helped me succeed in graduate school.

There are many friends and family who have supported me throughout the years. I am thankful for my parents, brother, sister, Wendy, Roland, Rob, and Karin for their continual support. I want to thank my friends Selasi, Maria, Nana, and Gigliola for supporting me in and out of lab. I am especially grateful to my best and closest friends Noelle, Radika, Shuyi, and Corey. Our dinners, texts, coffee dates, and deep conversations are cathartic and remind me there is more to life than just lab. All of them helped me get through the stresses of graduate school and life and gave me a sense of community.

Last, I want to thank my husband Tyler for listening to me discuss my work issues and being understanding about my experimental schedules, requiring him to looking after our family (Cielo and Titus) or even taking me to lab at odd hours. Tyler took care of me and reminded me that time-off is self-care. Without him, this journey would have been a lot more stressful and less fun.

CHAPTER 1: INTRODUCTION

Tuberculosis

Tuberculosis (TB) is a bacterial disease that usually infects the lungs and is caused by the intracellular bacterium *Mycobacterium tuberculosis*. TB disease is characterized by coughing, fever, night sweats, and weight loss (1). Globally, it is estimated that 2 billion people are infected with *Mtb*, and 5-10% of these individuals will develop TB disease (2, 3). If left untreated, 50% of patients will succumb to TB. *Mtb* killed an estimated 1.5 million people in 2020, ranking as the 13th leading cause of death and the second highest due to an infectious disease (3).

Tuberculosis treatment

Individuals infected with drug-susceptible TB are treated with a 6-month regimen of four antibiotics consisting of isoniazid, rifampicin, ethambutol, and pyrazinamide (3). However, the incidence of drug-resistant strains is increasing, including strains with more than monoresistance: multi-drug resistant TB (MDR-TB) is characterized by resistance to at least isoniazid and rifampicin, and extensively drug resistant TB (XDR-TB) by resistance to isoniazid, rifampicin, any fluoroquinolone, and to at least one of the injectable second-line drugs (1). Even more serious is the emergence of totally drug resistant TB (TDR-TB) that is resistant to all antibiotics. Globally, it is estimated that 5% of cases are multi-drug resistant TB (1). For individuals with multi-drug resistant TB, treatment times are longer, treatment more expensive, and associated with severe side-

effects. Despite having an effective drug regimen, the presence of drug-resistant infections demands continuous improvements to drug development.

Mycobacterium tuberculosis and host defenses

TB is caused by *Mycobacterium tuberculosis* (*Mtb*), a slow-growing intracellular bacterium that belongs to the *Mycobacterium tuberculosis* complex. One of the major features of *Mtb* is its unique cell envelope that contributes to its virulence and low drug permeability and is made up of the peptidoglycan layer, mycolic acid, and arabinogalactan (4, 5). *Mtb's* main virulence factors are its protein secretion systems, ESX1-5, and allow for full virulence of *Mtb (6-8)*.

During infection, *Mtb* encounters different host cells and thus defenses. *Mtb* initially infects the airway epithelial cells and phagocytes (macrophages, neutrophils, and dendritic cells). The infection progresses further as phagocytes are infected and the bacterium gains access to the lung interstitium, leading to recruitment of immune cells to the site of infection and forming the granuloma (1). During infection, the bacterium encounters a variety of host defenses. In phagocytic cells, *Mtb* encounters acid, oxidative, and nitric oxide stress (9, 10). However, *Mtb* counteracts these by blocking phagosome maturation and by resisting both oxidative and nitric oxide stress (11-17). Further host defenses include exposure to hypoxia, phosphate deprivation, and nutrient starvation, all which *Mtb* has evolved to survive (18-22).

Protein phosphorylation via two-component systems

Mtb is a successful pathogen in the face of many stresses and relies on sensing and responding to its environment. Protein phosphorylation is one mechanism used by all organisms to sense and respond to extracellular stress. The addition of a phosphate group to a protein can have a range of effects, including regulation of catalytic activity, affecting protein stability, changing protein-protein interactions, or changing sub-cellular localization (23). In bacteria, there are two main types of protein phosphorylation: one mediated by the two-component systems (TCSs) and the other by STPKs.

In bacteria, the two-component systems (TCSs) are the canonical protein kinase signaling systems. The TCSs consist of two proteins: a histidine kinase (HK) and a response regulator (RR). The former is a transmembrane protein kinase while the latter is a cytosolic transcription factor. In response to various cues, the sensor domain of the histidine kinase binds and “senses” specific signals, resulting in activation of the kinase domain. The activated kinase transfers phosphate to the response regulator to activate it, and the activated response regulator binds to DNA, typically to promote transcription of a specific set of genes, termed the regulon, resulting in a specific response to the stimulus. In *Mtb*, there are 12 complete pairs of TCSs, fewer TCSs compared to other bacteria relative to genome size (24, 25). The environmental triggers and downstream function of these TCSs have been studied extensively, including roles in nutrient acquisition, cell wall maintenance, and host adaptation (26-38). The TCSs have generally been the paradigm for protein phosphorylation systems regulating bacterial growth and pathogenesis.

Protein serine/threonine phosphorylation

Compared to the canonical bacterial TCSs, Ser/Thr /Tyr phosphorylation (*O*-phosphorylation) in bacteria was long understudied while in eukaryotes, *O*-phosphorylation have long been considered the mainstay of phosphosignaling. Eukaryotic *O*-phosphorylation shows a higher level of organization compared to the TCS that are typically linear and have one or few substrates (39). Importantly, therapeutic interventions often target STPKs – specifically in cancer – and STPKs are the second most targeted group among drug targets (40, 41). Altogether, *O*-phosphorylation is important for eukaryotic signaling but its role in bacteria is less clear.

O-phosphorylation in bacteria was initially studied more systematically about two decades ago. The first *O*-phosphoproteomes of various bacteria were probed and identified a small number of phosphoproteins and -sites: 105 phosphosites in *E coli*, 105 phosphosites in *B. subtilis*, 163 phosphosites in *S. pneumoniae*, and 55 phosphosites in *P. aeruginosa* (42-45). During this time, a pioneering study of the *O*-phosphoproteome in *Mtb* found that *Mtb* appeared to have one of the largest bacterial phosphoproteomes with over 500 phosphosites, which included sites only phosphorylated in different growth conditions (46). This study demonstrated that *Mtb* may have more *O*-phosphorylation compared to other bacteria and may rely more on *O*-phosphorylation for regulation. Recently, work in other bacteria have advanced techniques and improved conditions to detect more *O*-phosphorylated proteins, including one study in *E. coli* that identified 2,446 phosphosites on 861 proteins (47). The steady increase in *O*-phosphorylation identified in bacteria suggests that technical hurdles affected and may

yet affect our ability to comprehensively measure the phosphoproteome, including that of *Mtb*.

Serine/threonine protein kinases

Not only is bacterial protein *O*-phosphorylation less well defined than in eukaryotes, but it was also previously thought to be exclusive to eukaryotes. The bacterial STPKs are Hanks-type kinases with the 11 defining Hanks domains (48, 49) and were thus termed "eukaryotic-like". The number of STPKs encoded in bacteria varies widely: *E. coli* has 2 STPKs, *B. subtilis* has 4 STPKs, *H. pylori* has 1 STPK, and *S. typhi* has 3 STPKs (50), but others can have hundreds, such as *S. cellulosum* (51). In fact, the Global Ocean Sampling metagenomic study showed that at least in oceanic bacteria, the eukaryotic-like protein kinases are just as prevalent as histidine kinases (52). Many studies have shown a limited, but important role for STPKs in bacteria. Examples include regulation of quorum sensing in *Staphylococcus aureus* (53), activation of genes for virulence factors in *Streptococcus pyogenes* (54), and control of cell cycle and division in *Streptococcus pneumoniae* (55).

In *Mtb*, there are 11 STPKs, which are named PknA-PknL (there is no PknC). All STPKs, except PknG and PknK, are transmembrane proteins with an intracellular kinase domain and an extracellular sensor domain. Several STPKs have been linked to distinct environmental stresses, including PknB to low oxygen and PknD to osmotic stress (56, 57).These examples illustrate that these STPKs have similar roles to TCSs

as extracellular sensors. Interestingly, *Mtb* has an equal number of STPKs to TCSs and encodes a higher number of STPKs compared to other pathogenic bacteria.

STPK substrates and function

STPK function is defined by the protein(s) it phosphorylates, or its substrates. STPK substrates known in *Mtb* have been mostly identified by *in vitro* approaches. Most methods identified STPK-substrate pairs by *in vitro* phosphorylation assays using recombinant protein or peptides (46) Other studies relied on computational methods to identify specific substrate motifs in the proteome (58). From these studies, a small number of substrates for the *Mtb* STPKs were identified or predicted, and both have significant limitations. An even smaller number of phosphosites was subsequently shown to be functional in the bacterium (59-65).

A more faithful method to identify true cellular substrates is kinase perturbation in combination with MS. Two studies characterized pathways of PknB and its substrates CwlM and Lsr2, with the former showing phosphoregulation of peptidoglycan synthesis and the latter phosphoregulation of gene expression (66, 67). These examples illustrate the functional roles of STPKs in *Mtb*, but so far, have only focused on a subset of the STPKs. For a majority of the STPKs, there are no known cellular substrates or functions, and the vast majority of STPK-substrate interactions likely remains unknown.

Dissertation Aims

Our goal was to establish a systems-wide understanding of *Mtb O*-phosphorylation. We began by exploring the true size of the phosphoproteome and its general characteristics. In the deepest bacterial phosphoproteome to date, we show that *Mtb* has >5 times more phosphosites than previously reported. We then defined STPK function through the identification of substrates. Specifically, we used an approach that combines STPK LoF and GoF in combination with mass spectrometry-based phosphoproteomics to characterize *Mtb O*-phosphorylation. From these data, we created a resource that links thousands of individual phosphosites with their cognate STPK. With this resource, any *Mtb* STPK can now be linked to their ensemble of substrates, and any phosphosite can be linked to its upstream STPK. From this kinome-wide view emerged an STPK signaling system of a complexity that has previously only been described in eukaryotes (Chapter 3). We then characterized the role of phosphorylation in regulating transcription factors, the TCSs, and transcription itself. We identified a large effect of STPKs on gene expression and predicted many functional pathways from STPK to TF to gene. We validate one such pathway in detail (Chapter 4). Chapter 5 presents a summary of this dissertation and discusses future directions.

CHAPTER 2: MATERIALS AND METHODS

Mutant Generation and Culture Conditions

Loss-of-function (LoF) mutants were constructed by recombineering. The recombineering substrate was generated by amplifying 500 bp upstream and downstream of each STPK's coding sequence by PCR and cloning them 5' and 3' of a hygromycin-resistance cassette, respectively. A linear recombineering substrate was amplified by PCR, purified, and transformed into *Mtb* strain H37Rv carrying pNIT-ETc (68). Successful integration was confirmed using DNA sequencing. Gain-of-function (GoF) mutants were constructed using the episomal pDTCF plasmid encoding a tetracycline-inducible promoter, C-terminal FLAG tag, and a hygromycin-selectable marker. STPKs were amplified from *Mtb* strain H37Rv genomic DNA, cloned into the pDTCF plasmid 5' of the FLAG tag using Gibson assembly cloning, and confirmed with sequencing. Plasmids were transformed into *Mtb,* and expression was confirmed by anti-FLAG Western blot. The *pknB* CRISPRi-knockdown strain was constructed using a pJR965 plasmid encoding a tetracycline-inducible dCas9, a kanamycin-selectable marker, and a tetracycline-inducible sgRNA targeting the 5' coding sequence of *pknB* (69). STPK LoF and GoF were confirmed by DNA sequencing and selected reaction monitoring.

All strains were grown in Middlebrook 7H9 with 10% OADC and 0.2% glycerol at 37°C in rolling culture. For experiments involving STPK GoF, cultures were grown in 50 μg/ml hygromycin B. For experiments involving STPK CRISPRi knockdown, cultures were

grown in 30 µg/ml kanamycin. WT and STPK LoF strains were grown until stationary phase in triplicate. STPK GoF and CRISPRi-knockdown strains were grown in triplicate to OD_{600} 0.4-0.6, 100 ng/ml anhydrotetracycline added, and grown to stationary phase.

Lysate Production

Mtb cultures were pelleted at 4,000 x g for 5 min at 4°C, washed with phosphate buffered saline, and resuspended in lysis buffer (8 M Urea; 50 mM Tris pH 8.0; 75 mM NaCl; cOmplete protease inhibitor cocktail (Sigma 5892791001); 10 mM sodium fluoride (Sigma 919), 1 % phosphatase inhibitor cocktail 2 (Sigma, P 5726), 1% phosphatase inhibitor cocktail 3 (Sigma, P 0044)). Lysates were generated by bead-beating three times at 6 m/s for 30 sec with cooling on ice between steps. Samples were centrifuged at 12,000 x g for 1 min, and the supernatant sterilized by filtration. Protein concentrations were determined by Bradford assay and the lysates stored at -80°C until further processing.

Proteomic and phosphoproteomic sample processing

1 mg of protein from each sample in 1 ml lysis buffer was incubated for 15 min at 4°C. Protein concentrations were determined by BCA assay (ThermoFisher Scientific). Proteins were reduced with 5 mM dithiothreitol for 1 h at 37 °C, and then alkylated with 10 mM iodoacetamide for 45 min at 25°C in the dark. Samples were diluted 4-fold with 50 mM Tris-HCl, pH 8.0 to bring the urea concentration below 2 M for Lys-C digestion (Wako) at 1:50 enzyme-to-substrate ratio. After 2 hrs of Lys-C digestion at 25 °C, sequencing grade modified trypsin (Promega, V5117) was added to the samples at a

1:50 enzyme-to-substrate ratio, and incubated at 25°C for 14 hrs, and stopped by adding formic acid (Sigma-Aldrich, St. Louis, MO) to a 1% final concentration. Acidified samples were centrifuged for 15 min at 1,500 x g to remove precipitates, desalted on 200-mg tC18 SepPak (Waters, WAT054925) SPE, and concentrated down using a Speed-Vac concentrator. Final peptide concentration was determined via BCA Assay. Each sample was labeled with 10-plex tandem mass tags (TMT, ThermoFisher Scientific). TMT channel 131 was used for a pooled peptide mixture from all samples to normalize across experiments. Peptides (400 µg) were dissolved in 80 µl of 50 mM HEPES, pH 8.5 solution, and mixed with 400 µg of TMT reagent freshly dissolved in 20 µl anhydrous acetonitrile. 12 µL of 5% hydroxylamine was added to the samples and incubated for 15 min at RT to quench the reaction. Differentially labeled peptides were pooled, dried, reconstituted 3% acetonitrile, 0.1% formic acid solution, and desalted on tC18 SepPak SPE columns.

3.5 mg of 10-plex TMT labeled sample was separated on a reversed-phase Agilent Zorbax 300 Extend-C18 column (250 mm × 4.6 mm column containing 3.5-µm particles) using Agilent 1200 HPLC System. Solvent A was 4.5 mM ammonium formate, pH 10 in 2% acetonitrile and solvent B was 4.5 mM ammonium formate, pH 10 in 90% acetonitrile. The flow rate was 1 mL/min, and the injection volume was 900 µL. The LC gradient started with a linear increase of solvent B to 16% in 6 min, then linearly increased to 40% B in 60 min, 44% B in 4 min, 60% B in 5 min and then 14 min of 60% solvent B. A total of 96 fractions were collected into a 96-well plate. Fractions were concatenated into 24 fractions by combining 4 fractions that are 24 fractions apart (i.e.,

combining fractions #1, #25, #49, and #73; #2, #26, #50, and #74; and so on). For proteome analysis, 5% of each concatenated fraction was dried and re-suspended in 2% acetonitrile, 0.1% formic acid solution to a peptide concentration of 0.1 mg/mL for LC-MS/MS analysis. The remaining fractions (95%) were further concatenated into 12 fractions (i.e., by combining fractions #1 and #13; #3 and #15; and so on), dried, and subjected to immobilized metal affinity chromatography (IMAC) for phosphopeptide enrichment (70, 71).

Phosphopeptide enrichment

Fe3+-NTA-agarose beads were freshly prepared using the Ni-NTA Superflow agarose beads (QIAGEN, #30410) for phosphopeptide enrichment. For each of the 12 fractions (~200 ug per each fraction) and for each label-free sample (100 ug), peptides were reconstituted to 0.5 µg/µL in IMAC binding/wash buffer (80% acetonitrile, 0.1% trifluoroacetic acid), incubated with 10 µL Fe3+-NTA-agarose beads for 30 min at RT, and washed 2 times with 50 µL wash buffer and once with 50 µL 1% formic acid on the stage tip packed with 2 discs of Empore C18 material (Empore Octadecyl C18, 47 mm; CDS Analytical, 98-0604-0217-3). Phosphopeptides were eluted from the beads onto C18 using 70 µL of Elution Buffer (500 mM potassium phosphate buffer). 50% acetonitrile, 0.1% formic acid solution was used for phosphopeptide elution from C18 stage tips, dried, and reconstituted with 12 µL of 3% acetonitrile, 0.1% formic acid solution containing 0.01% DDM (n-Dodecyl-beta-Maltoside) for LC-MS/MS analysis.

Global and phosphoproteome analysis

Both TMT and label-free global and phosphopeptide enriched samples were subjected to a custom high mass accuracy LC-MS/MS system as previously described (72), where the LC component consisted of automated reversed-phase columns prepared in-house by slurry packing 3-µm Jupiter C18 (Phenomenex) into 35-cm x 360 µm o.d. x 75 µm o.d. fused silica (Polymicro Technologies Inc.). The MS component consisted of a Q Exactive HF Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific) outfitted with a custom electrospray ionization interface. Electrospray emitters were custom-made using 360 µm o.d. x 20 µm o.d. chemically etched fused silica capillary. Analysis of the phosphoproteome samples applied similar conditions used in the global proteome sample analysis, instead using a 2.2 kV spray voltage. All other instrument conditions were set as previously described (73).

Raw spectral data and analysis information is available via the MassIVE database for the public accession #MSV000088254. LC–MS/MS raw data were converted into dta files using Bioworks Cluster 3.2 (Thermo Fisher Scientific, Cambridge, MA, USA). The MSGF+ algorithm was used to search MS/MS spectra against the *Mtb* database (RefSeq H37Rv_uid57777_2014-08-14, 20198 entries). Search parameters included 20 ppm tolerance for precursor ion masses, +2.5 Da and -1.5 Da window on fragment ion mass tolerances, no limit on missed cleavages, partial tryptic search, no exclusion of contaminants, dynamic oxidation of methionine (15.9949 Da), static IAA alkylation on cysteine (57.0215 Da), and static TMT modification of lysine and N-termini (+144.1021 Da) for TMT analyses. The decoy database searching methodology (74, 75) was used

to control the false discovery rate at the unique peptide level to <0.01% and subsequent protein level to <0.1% (76). Quantification for TMT was based upon initially summing to the peptide (phospho) or protein (global) level for the sample specific peptide reporter ion intensities captured in each channel across all 24 or 12 analytical fractions. Final data for statistical analysis was scaling and central tendency normalization of each peptide or protein summed value across all observations within each TMT10 experiment to adjust for experiment-specific variability. Label-free quantification utilized MaxQuant-generated LFQ values (77) at the peptide (phospho) and protein (global) levels as previously described (78).

All collected protein and peptide abundance values were subjected to statistical comparison utilizing WT for comparison. Primary comparisons were based upon ANOVA analysis at a <0.005 p-value threshold coupled with a fold change criteria minimum of +/- 2 between abundances.

Selected reaction monitoring

SRM was performed on the panel of WT, GoF and LoF mutants as described above. Crude heavy peptides labeled with 13C/15N on C-terminal lysine and arginine were purchased from vivitide (Gardner, MA). Trypsin digested samples were processed as previously described (79). For each sample, the digested peptides were diluted to 0.25 µg/µL containing standards at a final concentration of 20 fmol/µL. and analyzed with a nanoACQUITY UPLC® system (Waters Cooperation, Milford, MA) coupled online to a TSQ Altis^{TM} triple quadrupole mass spectrometer (Thermo Fisher Scientific, San Jose,

CA). The LC-SRM platform was configured and utilized as previously described (79). The abundance of PknE and PknI in WT and LoF mutants were too low to be detectable through the standard LC-SRM analysis, so an extra PRISM procedure was applied to these samples to enrich for PknE and PknI peptides before LC-SRM analysis (80). SRM data acquired on the TSQ Altis™ were analyzed using Skyline-daily software (version 21.0.9.139) (81). Peak detection and integration were determined based on retention time and the relative SRM peak intensity ratios across multiple transitions between light peptide and heavy peptide standards (82). All data were manually inspected to ensure correct peak assignment and peak boundaries. The peak area ratios of endogenous light peptides and their heavy isotope-labeled internal standards (i.e., L/H peak area ratios) were automatically calculated, and the average peak area ratios from all the transitions were used for quantitative analysis of the samples. Correlation graphs were plotted for targets with more than one surrogate peptide to verify a strong correlation, and ultimately the peptide with the most sensitive response was selected for obtaining quantitative values.

RNA Isolation and RNA sequencing

Mtb cultures were pelleted at 4000 x g for 5 min at 4°C, resuspended in Trizol, transferred to a Lysing Matrix B tube, and lysed by bead-beating three times at 6 m/s for 30 sec with cooling on ice between steps. Samples were centrifuged at 21,000 x g for 1 min, the supernatant was transferred to a Heavy Phase Lock Gel tube containing 300 µl chloroform, inverted for 2 minutes, and then centrifuged at 21,000 x g for 5 min. RNA in the aqueous phase was precipitated using 300 µl isopropanol and 300 µl high-salt

solution (0.8 M Na citrate, 1.2 M NaCl). Total RNA was purified using a RNeasy kit with one on-column DNase treatment and quantified using a Nanodrop. mRNA was enriched by depleting rRNA from samples using the Ribo-Zero rRNA removal (bacteria) magnetic kit (Illumina), and then was prepared for Illumina sequencing using the NEBNext Ultra RNA Library Prep Kit for Illumina according to the manufacturer's instructions in combination with AMPure XP reagent for clean-up of adaptor-ligated DNA. Each replicate was barcoded using the NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1). 30-40 libraries were multiplexed, quantified using the KAPA qPCR quantification kit, and sequenced at the University of Washington Northwest Genomics Center with the Illumina NextSeq 500 High Output v2 Kit. Read alignment was performed using the custom processing pipeline that uses Bowtie 2 utilities, available at https://github.com/robertdouglasmorrison/DuffyNGS and https://github.com/robertdouglasmorrison/DuffyTools.

Phosphorylation Motif Analysis

The custom background peptidome was generated by using the H37Rv proteome from UniProt.org (Proteome ID: UP000001584, modification date 12/1/2019). A custom R script was used to partition the proteome into unique, 11-mer peptides. Sequences were further subdivided into groups with either serine, threonine, or tyrosine as central residues. The foreground peptidome was made using the differentially modulated phosphosites for each kinase. R script was used to make two groups of 11-mer peptides: all phosphosites identified in the WT phosphoproteome (basal phosphoproteome) and STPK-modulated phosphosites. For motif analysis, four

separate analyses were performed for each kinase and the basal phosphoproteome: a site-centered or a Ser/Thr/Tyr-centered motif search. Respective foreground sequence and appropriate background peptidomes were uploaded to the pLogo web application for analysis using default parameters.

STPK auto-phosphorylation

PknB-PknL kinase domains were PCR- amplified from *Mtb* strain H37Rv genomic DNA, cloned into inducible pET28b plasmid 3' of the His tag using Gibson assembly cloning, confirmed with sequencing, and transformed into BL21 *E. coli*. BL21 cultures were grown to 0.4-0.6 OD, induced with 250 µM IPTG, and grown at 20°C for 18 - 24 hours. Cultures were pelleted at 4000 x g for 20 min, resuspended in lysis buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 20 mM imidazole), and then lysed using sonication. Lysates were centrifuged at 35,000 x g for 30 min, initially purified using Ni-NTA affinity chromatography beads, further purified using size exclusion chromatography, and then stored in size exclusion buffer (20 mM Tris-HCl, 150 mM NaCl, 5% glycerol, pH 7.4). For auto-phosphorylation, 1 µM STPK was added to reaction buffer (50 µM MnCl2, 50 µM MgCl2, 1 mM DTT, 50 mM Tris pH 8.0, 75 mM NaCl) , initiated with 50 µM ATP and 15 µCi ATP [γ-32P], and incubated at 37°C for 30 min. Reactions were terminated using 4x loading buffer with β-mercaptoethanol, boiled at 95°C for 5 min, and then SDS-PAGE was performed using a 4-15% Bis-Tris Plus gel run at 175V for 30 min. Gels were stained with Coomassie Brilliant Blue for 1 hr, de-stained in water overnight, and dried at 80°C for 3 hrs. Phosphor screens were exposed to the gel overnight, imaged using a Sapphire biomolecular imager, and quantified with ImageJ.

Transcription Factor Phosphorylation

Recombinant Rv2359 (Zur) was cloned, expressed, and purified using the protocols described above. Protein was desalted with PD-10 desalting columns and eluted using size exclusion chromatography buffer (20 mM Tris-HCl, 150 mM NaCl, 5% glycerol, pH 7.4). For the phosphorylation assay, 10 μM Rv2359 was combined with 1 μM STPK in 20 μl reaction buffer (50 μM MnCl2, 50 μM MgCl2, 1 mM DTT, 50 mM Tris pH 8.0, 75 mM NaCl), initiated with 50 μM ATP and 15 μCi ATP [γ-32P], and incubated at 37°C for 30 min. Reactions were terminated using 4x loading buffer with β-mercaptoethanol, boiled at 95°C for 5 min, and then SDS-PAGE was performed using a 4-15% Bis-Tris Plus gel run at 175V for 30 min. Gels were stained with Coomassie Brilliant Blue for 1 hr, de-stained in water overnight, and dried at 80°C for 3 hrs. Phosphor screens were exposed to the gel overnight, imaged using a Sapphire biomolecular imager, and quantified with ImageJ.

Zur Mutagenesis

QuikChange site-directed mutagenesis was used to make the Zur phosphomimetic (Thr67Asp) and phosphoablative (Thr67Ala) mutants (83). PCR products were gel extracted, DpnI-digested, transformed into competent *E. coli*, and confirmed with sequencing. Recombinant Zur mutant proteins were made using the protocols described above.

Electrophoretic Mobility Shift Assay (EMSA)

Cy5-labeled DNA probes were made by combining three oligos (Integrated DNA Technologies) each to a final 100 µM concentration in dsDNA annealing buffer (10 mM Tris-HCl-pH 7.5, 100 mM NaCl, 1 mM EDTA). Oligo 1 is a ChIP-identified DNA binding region of Rv2359 (84). Oligo 2 is the reverse complement of Oligo 1with a reverse complement of the Oligo 3 12-mer. Oligo 3 is a 12-mer with Cy5 covalently coupled to the 5' end. The oligo mixture was vortexed and heated to 95°C for 10 min and cooled to RT over 3-4 hrs protected from light. For EMSA, recombinant Rv2359 WT and mutant protein were combined with 2 µM final DNA probe in 20 µl reaction buffer (20 mM Tris-HCl pH 8.0, 75 mM NaCl, 1 mM DTT, 50 µM MgCl2, 50 µM MnCl2, 5% glycerol, 1% zinc sulfate, 50 µg/mL bovine serum albumin, 50 µg/mL salmon sperm DNA) and incubated for 30 min at RT, protected from light. After incubation, 4x native sample loading buffer was added to each sample and loaded on a native 10% Tris-Glycine gel. The gel was maintained on ice and ran at a constant 100 volts for 2 hrs. The gel was rinsed with water, stored in phosphate buffered saline, visualized on a Li-cor Odyssey scanner. Bands were quantified using the Odyssey software.

NarS auto-phosphorylation and NarL phosphotransfer

Recombinant NarS cytoplasmic domain (AA 220-425) was cloned into pET28b using the protocols described above. QuikChange site-directed mutagenesis protocol was used to generate NarS phosphomimetic (Thr380Asp and Thr380Glu) and phosphoablative (Thr380Ala) mutants. PCR products were gel extracted, DpnI-digested, transformed into competent *E. coli*, and confirmed with sequencing. Recombinant NarS

WT, Thr380Asp, Thr380Glu, and Thr380Ala protein were expressed and purified using the protocols described above. For auto-phosphorylation, 20 μM NarS WT and mutants were individually added to 20 μl reaction buffer (50 mM Tris-HCl pH 8.0, 50 mM KCl, 10 mM MgCl2), initiated with 50 μM ATP and 15 μCi ATP [γ-32P], and incubated at 30°C for 10-30 min. Reactions were terminated using 4x loading buffer with β-mercaptoethanol, and then SDS-PAGE was performed using a 4-15% Bis-Tris Plus gel run at 175V for 30 min. Gels were stained with Coomassie Brilliant Blue for 1 hr, de-stained in water overnight, and dried at 80°C for 3 hrs. Phosphor screens were exposed to the gel overnight, imaged using a Sapphire biomolecular imager, and quantified with ImageJ.

Recombinant NarL was cloned, expressed, and purified using the protocols described above. NarS was auto-phosphorylated as described above, and then phosphotransfer was performed by adding 20 μM NarL, and incubating for 2-10 mins at 30°C. Reactions were terminated using 4x loading buffer with β-mercaptoethanol, and then SDS-PAGE was performed using a 4-15% Bis-Tris Plus gel run at 130V for 2 hrs. Gels were stained with Coomassie Brilliant Blue for 1 hr, de-stained in water overnight, and dried at 80°C for 3 hrs. Phosphor screens were exposed to the gel overnight, imaged using a Sapphire biomolecular imager, and quantified with ImageJ.

CHAPTER 3: EXTENSIVE *O*-PHOSPHORYLATION IN *MTB*

Introduction

STPK function is defined by its substrates. Over the last two decades, studies used a variety of experiments to identify STPK-substrate interactions. Starting with candidate interactions based on co-expression of the STPK with genes in the same operon or known STPK interactions such as those with FHA-domain-containing proteins, many studies used *in vitro* kinase assays on *Mtb* peptides or recombinant protein with varying success (46). Another study used a *Mtb* proteome microarray to globally identify protein-protein interactions for all STPKs, finding 492 binding proteins and over 1,000 interactions (85). Proteome arrays can successfully candidate STPK substrates, but interactions need validation, and many STPK-substrate interactions are transient. Altogether, these *in vitro* studies can successfully identify some STPK substrates, but it is unclear whether most of these interactions represent true cellular interactions.

Recent work in *Mtb* has focused on identifying STPK substrates in a cellular context, using a combination of phosphoproteomics and kinase loss-of-function mutants or chemical inhibition. For example, one group identified PknB substrates by comparing PknB-depleted *Mtb* to WT in osmoprotective media and showed that PknB regulates peptidoglycan synthesis via CwlM phosphorylation (66). While recent work has embraced this method of identifying substrates in *Mtb*, they still only link a minority of substrates to cognate STPKs and are only focused on individual STPKs. Importantly, these studies overlook the fundamental context of kinase activation (86). These

methods begin to identify more physiologically relevant STPK-substrate pairs in *Mtb*, but there likely remain a significant number of cellular interactions to be identified in other STPKs and during STPK activation.

Studies in eukaryotes have pioneered new cellular approaches to comprehensively identifying kinase substrates. A yeast phosphoproteomics study comprehensively inactivated kinases by knockout or chemically inhibition and used mass spectrometry to measure differences in phosphosite abundance compared to WT. This study defined significant phosphorylation changes as kinase-substrate interactions and identified interactions between nearly 100 kinases and more than 1,000 phosphoproteins (87). This approach demonstrated that kinases can be studied systematically and showed the global relationship between kinases and their substrates by defining individual and co-regulated kinase-substrate interactions. However, these studies do not identify substrates during kinase activation. Since *Mtb* has a fraction of the kinases compared to yeast, it is feasible to consider a similar approach in *Mtb*.

Further, the similarity between eukaryotic and bacterial STPKs in sequence and structure has made the vastly better studied eukaryotic STPKs a template for understanding STPK signaling in bacteria. Studies of eukaryotic STPKs identified interactions between STPKs and extensive crosstalk that can modulate pathway output and form branched signaling networks with multiple substrates, in contrast to the bacterial TCS that only phosphorylate their specific response regulator. Bacteria have between none to hundreds of STPKs. *S. cellulosum*, for example, codes for 317 STPKs

(51), a number similar to eukaryotic organisms that may imply a similar network organization. A comprehensive analysis of hundreds of STPKs is challenging, and a systems-wide analysis of a larger bacterial kinome is currently lacking. *Mtb*, however, expresses eleven STPKs, a kinome that is small enough to be experimentally tractable yet large enough to potentially give rise to a network of STPK signaling. In fact, one study identified a number of *in vitro* STPK-STPK interactions, suggesting a more complex STPK phosphorylation network structure (88). Overall, the presence and organization of a higher-order bacterial STPK phosphorylation system in bacteria remains unknown.

In this chapter, we addressed the question about the presence and potential organization of an STPK network in *Mtb* by comprehensively identifying STPK-substrate interactions by combining kinase perturbation with mass spectrometry-based phosphoproteomics. The differences in STPK LoF and GoF mutant phosphoproteomes compared to WT that are indicative of substrates are hypo-phosphorylation in the LoF and/or hyper-phosphorylation in GoF mutants. By using highly sensitive MS instrumentation, sample fractionation, TMT and label-free approaches, and 20 different STPK mutant strains, we generated a deep *O*-phosphoproteome. We identified multiple substrates for each STPK and begin to comprehensively identify STPK function. Interestingly, we identified a system of interactions between STPKs that resembles a eukaryotic kinase signaling network in its complexity. In addition, this phosphoproteomic dataset serves as a resource for assigning phosphorylated proteins to their cognate

STPKs. With this resource, any *Mtb* STPK can now be linked to their spectrum of substrates, and any phosphoprotein can be linked to its upstream STPK.

Results

An STPK gain- and loss-of-function mutant panel

We sought to comprehensively characterize the *Mtb* phosphoproteome and test the effects of individual STPK perturbation by measuring changes in phosphorylation in STPK mutant strains. In *Mtb*, 10 of the 11 STPKs are situated in or at the cell membrane, and nine STPKs have an extracellular sensor domain that extends into the periplasmic space. We anticipated that several *Mtb* STPKs are inactive in axenic culture and require activation through signals that are likely absent in culture. The loss of such an inactive enzyme by knockout would not produce measurable changes. STPKs, also, are typically activated by binding to ligands that promote oligomerization, and this mode of activation appears to be generally conserved in *Mtb* (89*)*. Further, the physiologic ligands of the STPKs are known only for PknB, precluding activation of most STPKs by their natural ligands, so we used overexpression to force STPK activation in the absence of ligand (90). Thus, we developed a reciprocal perturbation approach that combines kinase LoF (knockout) and GoF (overexpression) to reveal substrates for all STPKs irrespective of their activation state.

All mutant strains were generated in the *Mtb* H37Rv background. GoF strains were engineered by inserting a plasmid expressing the STPK coding region in frame with a

C-terminal FLAG tag under control of the inducible tetracycline promoter. For LoF mutants, we deleted all STPK genes, except for the *in vitro* essential PknA and PknB, by inserting a hygromycin cassette in place of the STPK coding region using recombineering (68). To create a PknB LoF mutant, we used CRISPRi to reduce PknB expression. We were unable to obtain a PknA GoF strain. The *pknA* gene caused toxicity in *E. coli* and precluded subcloning. A catalytically inactive PknA mutant could be readily cloned and expressed in *E. coli* and *Mtb*, demonstrating that the specific kinase activity was responsible for toxicity (data not shown). Also, a study of the phosphoproteomic consequences of PknA deletion was recently presented, prompting us to exclude PknA from our study (72). We grew all cultures to early stationary phase, as slowed replication under nutrient limitation arguably mimics restrictive *in vivo* growth conditions more faithfully than exponential growth, and because the largest degree of *O*-phosphorylation in *E. coli* was reported in stationary phase (91). We confirmed STPK knockout and induction by quantitative, targeted mass spectrometry using selected reaction monitoring (SRM) (**Figure 3-1A**). All STPKs were absent or reduced to below the level of detection in the LoF strains, confirming knockout or knockdown, and all GoF mutants showed robust induction. In total, we generated 20 STPK mutant strains, a complete LoF/GoF panel except for the PknA mutants.

A deep *Mtb* phosphoproteome

To define a comprehensive *Mtb O*-phosphoproteome, we analyzed WT and STPK mutant strains by mass spectrometry-based phosphoproteomics. Previous *Mtb* phosphoproteomics studies used label-free mass spectrometry to identify

phosphoproteome changes. However, we used both label-free and labeled approaches to test our 20 mutant strain panel and WT. The latter approach incorporated a tandem mass tag (TMT) to label individual samples and multiplexed prior to mass spectrometry analysis. Interestingly, the label-free approach identified more total phosphosites, but the TMT produced more consistent quantitation. To compile the most comprehensive phosphoproteome, we combined the label-free and TMT data since they both identified unique, high-confidence phosphosites. We analyzed total protein abundance and phosphorylation for each of the 20 mutant strains and WT, in biological triplicate. Together, these data produced >150 phosphoproteomes and >105 million MS^2 spectra in total. To obtain the most confident and accurate phosphosites, we only considered phosphosites that were detected in at least one of the three replicates and had a MaxQuant probability score >0.99 (label-free) or an Ascore >19 (TMT), both scores representing a 99% confidence in phosphosite assignment.

In our global MS experiments, we detected 3,323 of the ~4,200 predicted *Mtb* proteins, detecting high coverage compared to previous *Mtb* proteogenomic studies and in light that only 80% of *Mtb* genes are expressed at a time (92, 93). Using the phosphosite identification criteria, we identified 10,244 unique phosphosites using TMT, 8,327 unique phosphosites using label-free, and 4,119 phosphosites in both experiments **(Figure3-1C)**. In total, we identified 14,452 unique phosphosites on 2,617 *Mtb* proteins (**Figure3-1B**), accounting for 87% of MS-detected *Mtb* proteins. These data showed extensive *O*-phosphorylation in *Mtb* that is comparable to the highest level of phosphorylation detected in eukaryotes, where 50-70% of proteins are phosphorylated.

We identified five times more phosphosites than all previously published *Mtb* phosphoproteome studies combined (**Figure3-1B**). Importantly, we identified the largest bacterial phosphoproteome to date.

In total, we identified 7,533 pThr, 5,816 pSer, and 1,103 pTyr sites. The distribution of site localization scores with both methods showed that most sites could be identified with high confidence, with ~50% of label-free and TMT assignments reaching >99% confidence and over 60% reaching 90% confidence (**Figure 3-1D**). We observed that the distribution of pSer, pThr, and pTyr detected in our study was 40%, 52%, and 8%, and this differed from that of eukaryotic or *E. coli* phosphoproteomes, both of which have a higher proportion of pSer than *Mtb* (**Figure 3-1E**). The striking features of the *Mtb* phosphoproteome was the predominant pThr compared to both eukaryotes and *E. coli*, and higher pTyr compared to eukaryotes. Specifically, we detected more pTyr sites than previous *Mtb* studies, and total *Mtb* pTyr is >10-fold higher than in humans, indicating a more prominent role of pTyr in *Mtb* (94, 95).

Multi-site phosphorylation in *Mtb* proteins

Given that there is > 5-fold more phosphosites than proteins, we next determined the number of phosphosite per protein (**Figure 3-1F**). We found that while a large group of proteins had a single phosphosite (>500), most proteins contained two or more phosphosites, with 413 proteins with two sites and 1,330 proteins that have between 3 and 10 sites. We also observed a small group of proteins with >10 sites, with one protein containing >80 phosphorylation sites. These highly multi-phosphorylated

proteins included the STPK PknK, forkhead-associated domain-containing protein FhaA, fatty acid synthetase Fas, DNA topoisomerase TopA, and NAD-dependent glutamate dehydrogenase Gdh (**Figure 3-1F, inset**). Highly phosphorylated proteins could suggest a function outside of the canonical on-off switch typically associated with phosphorylation and may imply a physiochemical property change due to the large accumulation of negative charges in a cellular sub-compartment.

To test if phosphorylation correlates with *in vitro* essentiality, we determined the number of phosphoproteins that are essential as defined by *in vitro* transposon mutagenesis in relation to their number of phosphosites (96) (**Figure 3-1G**). The phosphoproteome overall is enriched ~1.5 fold in essential genes, but when we analyzed the essential proteins across the multi-phosphorylated proteins, we saw a strong correlation between essentiality and total phosphosites on an individual protein. For example, 36% of proteins with more than nine phosphosites were essential but ~88% of those with 31 - 40 sites were essential ($p < 0.001$), compared to ~20% essential genes in the genome overall.

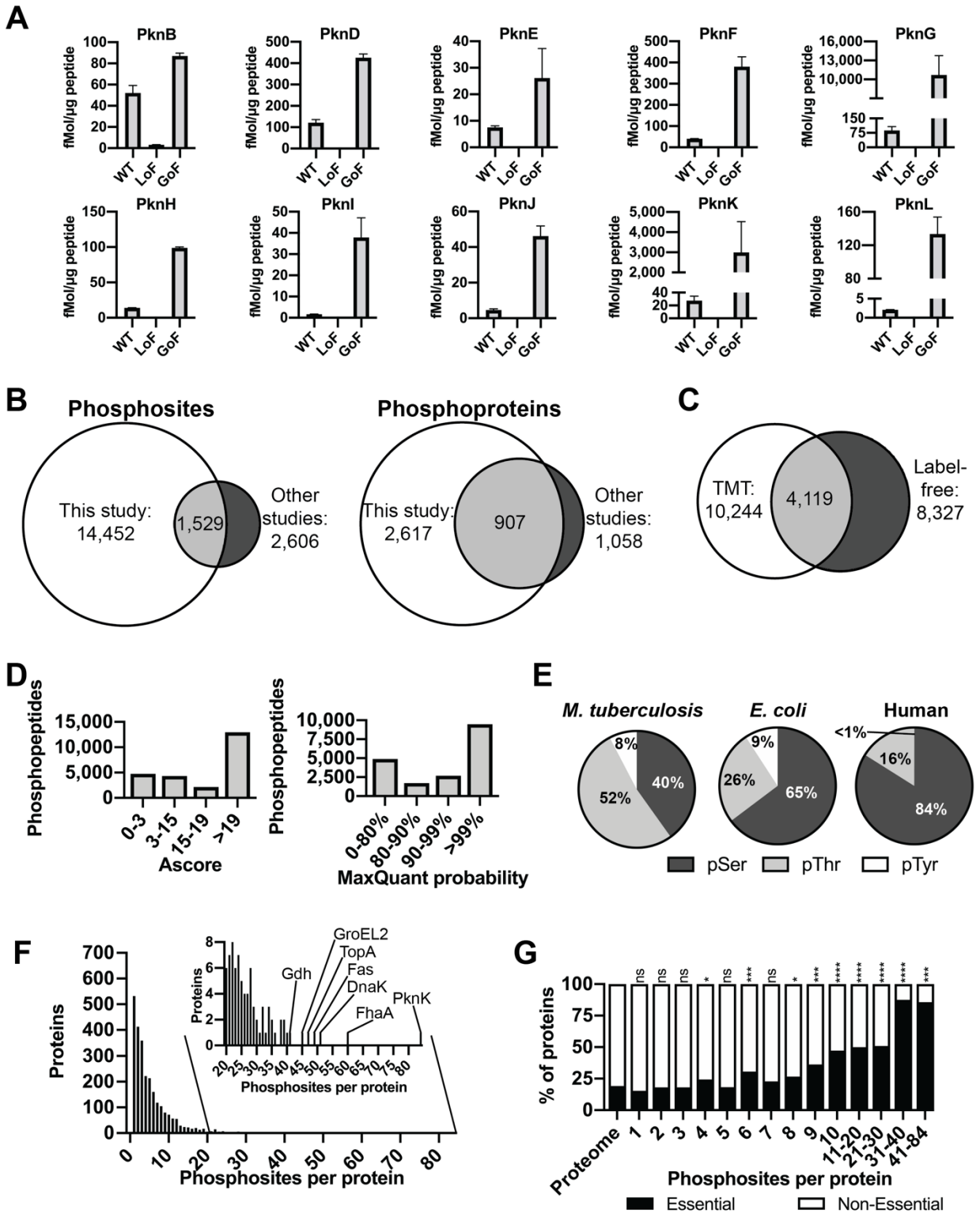

A
PknB
PknD
PknE
PknF
PknG
PknH
PknI
PknJ
PknK
PknL
fMol/µg peptide
WT
LoF
GoF
B
Phosphosites
This study: 14,452
1,529
Other studies: 2,606
Phosphoproteins
This study: 2,617
907
Other studies: 1,058
C
TMT: 10,244
4,119
Label-free: 8,327
D
Phosphopeptides
Ascore
0-3
3-15
15-19
>19
MaxQuant probability
0-80%
80-90%
90-99%
>99%
E
M. tuberculosis
E. coli
Human
8%
40%
52%
9%
26%
65%
<1%
16%
84%
pSer
pThr
pTyr
F
Proteins
Phosphosites per protein
GroEL2
TopA
Gdh
Fas
DnaK
PknK
FhaA
G
% of proteins
ns
*
***
****
Proteome
Phosphosites per protein
Essential
Non-Essential

Figure 3- 1. A deep *Mtb* phosphoproteome.

(A) Absolute quantitation of STPK protein levels in LoF and GoF mutant strains by selected reaction monitoring. PknB LoF is a knockdown strain, all other LoF strains are knockouts.
(B) Comparison of the number of phosphosites and -proteins previously published and identified in this study. Phosphosites identified in this study met an Ascore >19 (TMT) or MaxQuant probability >0.99.
(C) Number of unique phosphosites identified by TMT and label-free approaches. All phosphosites met an Ascore >19 (TMT) or MaxQuant probability >0.99.
(D) Distribution of Ascores (for TMT data) and MaxQuant (for label-free data) probability for phosphosite localization. An Ascore of 19 corresponds to 99% probability of correct localization.
(E) Relative share of pSer, pThr, and pTyr sites in *Mtb*, *E. coli*, and human cells.
(F) Number of phosphorylation sites per protein and highly phosphorylated proteins with >20 phosphosites (inset).
(G) Share of phosphorylated essential gene products stratified by number of phosphosites. Essentiality was based on transposon mutagenesis (96).

STPK phosphorylation

STPKs are known to be regulated by phosphorylation (89). In the total phosphosite dataset, we identified all STPKs as phosphorylated (**Table 1**). PknK had the most phosphosites with 84 and PknI had the least with 4. Phosphosites were mainly distributed across the cytoplasmic domain, including the kinase and juxta-membrane domains. The transmembrane-containing STPKs (PknB, PknD, PknE, PknF, PknH, PknJ) also had some phosphosites in the extracellular domain.

Using previous phosphoproteomics studies, we detected 91% (64 of 70) of previously identified STPK phosphosites (46, 72, 97-99). Interestingly, no other study identified phosphosites on PknI or PknK, but we identified four significant phosphosites on PknI and 84 on PknK. One study had previously identified crucial residues in the activation loop for nine STPKs using recombinant protein, and in our dataset, we identified these phosphosites for six of the nine STPKs (88). Altogether, we validated other phosphosites identified in previous phosphoproteomics studies and discovered new phosphosites that may be important for STPK regulation.

STPK	Rv Number	Num of pSites	pSites
PknA	Rv0015c	16	S10, T125, T150, T152, T174, T21, T224, T252, S299, T301, S309, T313, T323, T332, S46, S50
PknB	Rv0014c	18	T111, T171, T173, T179, Y182, S244, T265, T289, T294, S305, T309, T320, T352, T376, S51, Y53, T612, T621
PknD	Rv0931c	31	S10, Y107, S164, T169, T171, T173, T177, Y180, T187, S2, S218, S235, S264, S279, T295, T299, T303, S306, S308, S310, T317, T321, S328, S332, T334, S343, T35, S355, S365, S51, T633
PknE	Rv1743	28	T11, S125, Y16, T170, T175, T178, Y180, S188, S190, T193, T263, T277, T284, S290, T300, S304, T313, S326, T36, T4, T425, S448, S48, T50, T510, S517, S52, T523
PknF	Rv1746	16	T178, T181, S20, Y27, T287, S290, T378, S380, T382, T398, T458, T461, S48, Y53, S66, T8
PknG	Rv0410c	15	T14, T23, S298, Y307, T32, T362, S412, T55, S643, T657, S720, S723, Y735, T736, T95
PknH	Rv1266c	26	S103, S11, Y16, T170, T174, T178, S188, T193, S219, S291, T295, T307, T324, S335, T352, T36, S361, Y376, T48, S52, T537, T549, T551, T574, Y82, T99
PknI	Rv2914c	4	T207, S230, S320, S44
PknJ	Rv2088	7	T168, T27, S380, T382, S383, S39, S6
PknK	Rv3080c	84	T1010, T1019, T1021, S1028, S1034, S1038, T1058, T1093, S110, S1101, T113, T126, S128, T158, T166, T2, T219, Y224, S228, T240, S241, T274, S292, T319, T323, T325, T328, T331, Y336, S339, T342, T347, S349, T352, S370, S375, T376, Y41, T421, S425, Y440, T443, S444, S467, S47, S514, T521, T545, T546, S561, T569, S58, T59, T593, T611, S617, T623, T642, S679, Y687, T713, S719, T722, T723, T733, S734, T760, T768, T779, T825, T829, T87, T888, T894, T9, S900, S908, T917, S930, S935, Y942, S968, T969, T973
PknL	Rv2176	24	S151, S17, S186, S239, T257, S27, S295, T30, S300, Y304, S306, T309, S31, T323, S334, S338, Y34, S340, T348, S52, Y54, T6, T62, Y83

Table 1: Phosphosites identified on the STPKs.

Rows show each individual kinase, its Rv number, the total number of phosphosites identified per kinase, and the total residues phosphorylated per kinase. Red text indicates phosphosites regulated by other STPKs.

The *Mtb* kinase-substrate relationships

Next, we used our phosphoproteomic data to globally identify substrates for each STPK. Using our panel of STPK mutants, we defined substrates of a given STPK as either hypophosphorylated in that kinase's LoF or hyperphosphorylated in the GoF strain. Proteins with phosphorylation changes in the opposite direction were considered indirect substrates. We compared the abundance of each phosphosite in every STPK mutant to that of WT, calculating both fold-change and p-values. We used the TMT data for this analysis because TMT showed higher consistency between replicates and less missing values compared to the label-free data. We defined a significant change as one detected in all three replicates for WT and the STPK mutant, changed by at least 2-fold compared to WT, and with a p-value <0.005. In addition, we identified "on/off" sites as those with no detected peptides in one set of replicates, but for these sites, we were unable to calculate p-values.

In total, we identified 3,503 significant, unique changes in the phosphoproteome across all mutants that accounted for 24% of all phosphosites (**Figure 3-2**). For a majority of the STPK mutants, we observed the overall shift in the phosphoproteome in response to kinase perturbation were as expected, with GoF leading to increases in phosphorylation and LoF to decreases. Together, we detected 296 and 3,454 changes because of LoF and GoF, respectively (**Figure 3-2**). The STPKs varied in the number of substrates ranging from 15 - 968, suggesting varying signaling capacity of the different STPKs (**Figure 3-2, Figure 3-3A and C**).

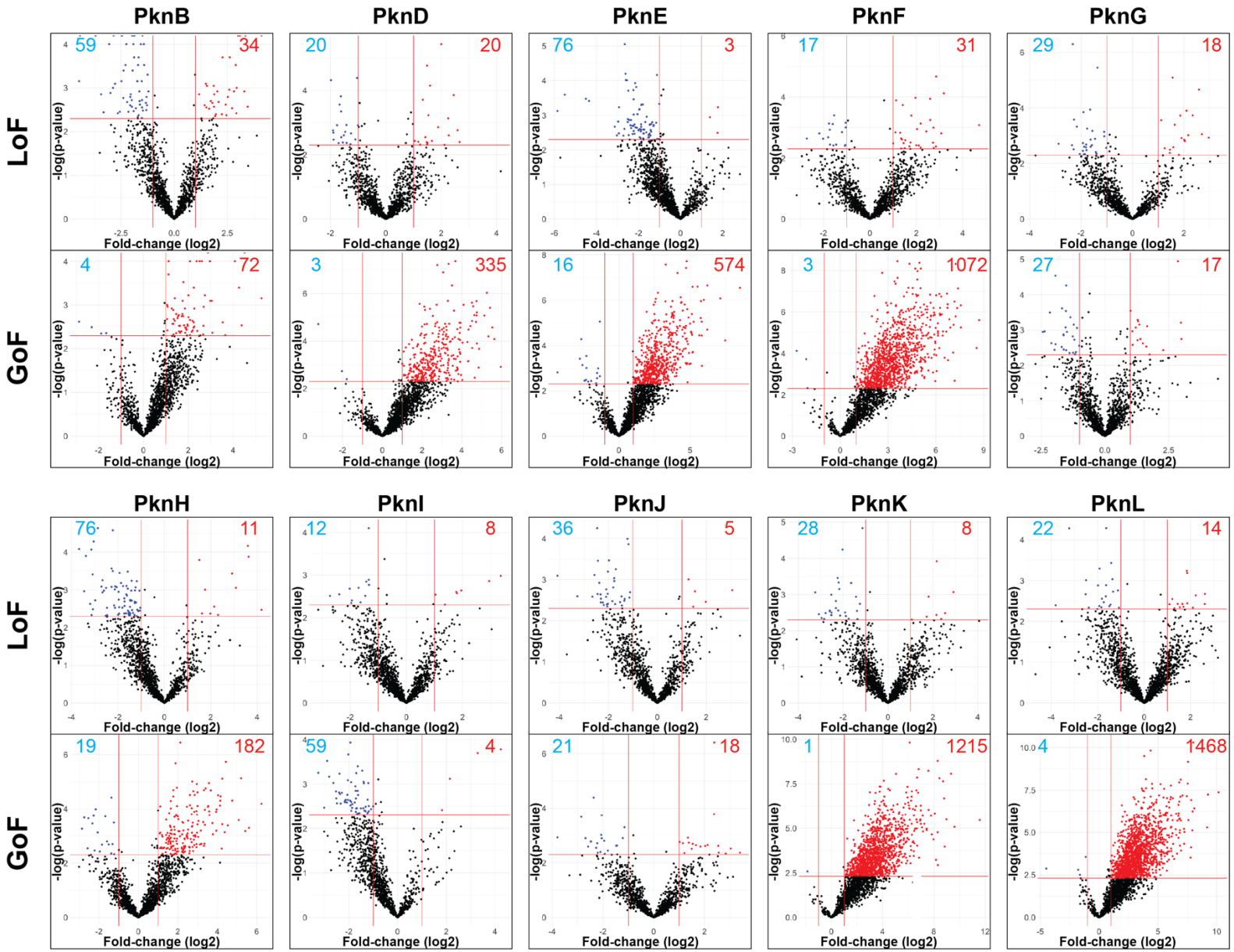
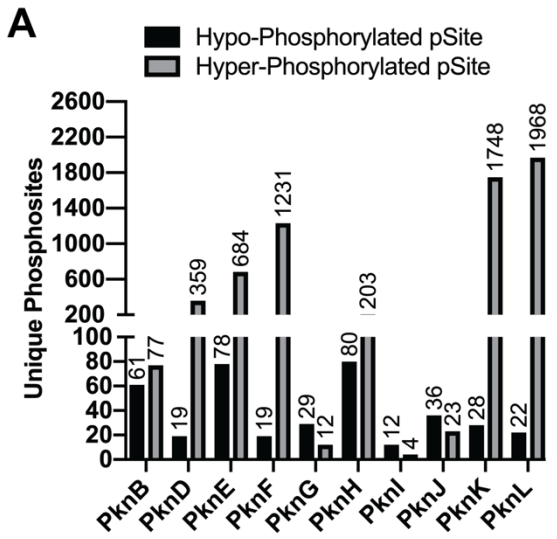
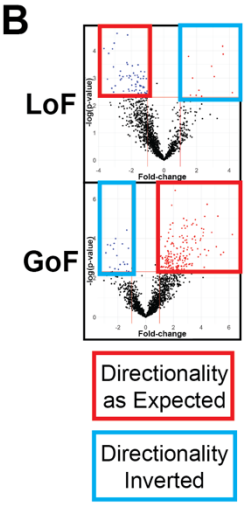

PknB
PknD
PknE
PknF
PknG
PknH
PknI
PknJ
PknK
PknL
LoF
GoF
-log(p-value)
Fold-change (log2)
59
34
20
20
76
3
17
31
29
18
4
72
3
335
16
574
3
1072
27
17
76
11
12
8
36
5
28
8
22
14
19
182
59
4
21
18
1
1215
4
1468

Figure 3- 2: STPK perturbation defines kinase-substrate interactions.

Magnitude and direction of changes in phosphorylation sites in LoF (Rows 1 and 3) and GoF strains (Rows 2 and 4). Volcano plots show log2-fold changes in phosphorylation levels by phosphosite (X-axis) versus the -log p value of the change (y-axis). Horizontal and vertical lines show significance cutoffs applied to the data for further analysis (>2-fold change in abundance, $p<0.005$). Significantly changing phosphosites are shown in red and blue. The STPKs themselves were removed as they showed the largest changes and distorted the plots.

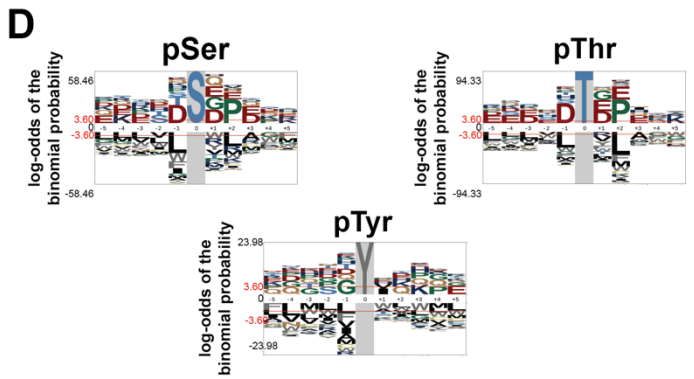


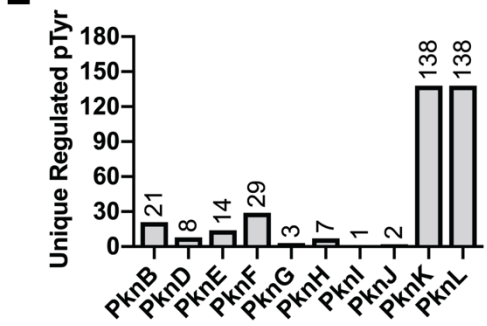


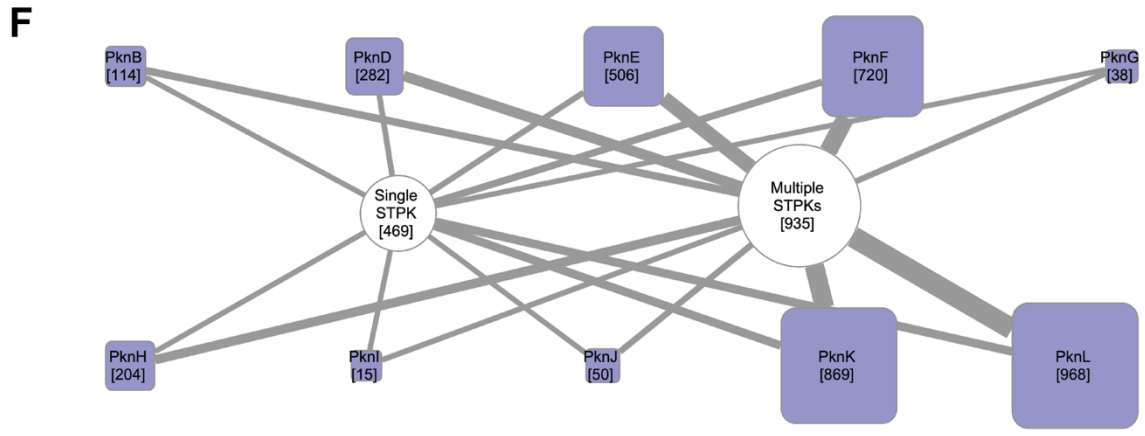
C

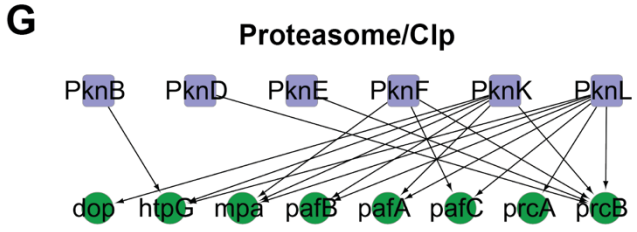
Regulated pSites by STPK

STPK	Direct pSites (Proteins)	Indirect pSites (Proteins)
PknB	136 (114)	40 (32)
PknD	376 (282)	24 (23)
PknE	742 (506)	47 (41)
PknF	1241 (720)	76 (69)
PknG	41 (38)	50 (41)
PknH	273 (204)	37 (31)
PknI	16 (15)	69 (60)
PknJ	58 (50)	28 (26)
PknK	1760 (869)	148 (131)
PknL	1978 (968)	105 (96)









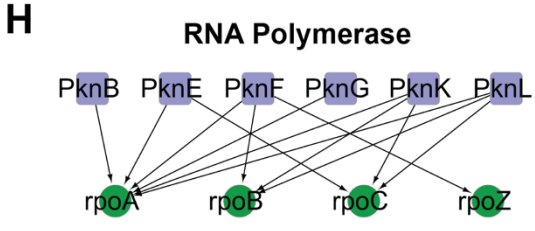


Figure 3- 3: *Mtb* O-phosphorylation forms a network.

(A) Effects of STPK perturbation on the abundance of individual phosphosites by STPK, compared to WT. All regulated phosphosites met the following criteria: > 2-fold change in GoF and/or < -2-fold change in LoF, Ascore > 19, and p-value <0.005.
(B) Directionality of phosphorylation changes in response to STPK perturbation. The illustration shows assignment of direct (hypophosphorylation in LoF, hyperphosphorylation in GoF), and indirect phosphorylation events with inverted directionality.
(C) Putative direct and indirect phosphosites (phosphoproteins) by STPK.
(D) Preferred sequence motifs of all STPKs around pSer, pThr, and pTyr sites.
(E) Changes in Tyr phosphorylation sites upon STPK perturbation. The numbers given above the STPK mutant strain is from LoF and GoF mutants combined. All regulated phosphosites met the following criteria: > 2-fold change in GoF and/or < -2-fold change in LoF, Ascore > 19, and p-value <0.005.
(F) A global view of the STPKs' substrates and overlap in substrate phosphorylation. The purple squares represent the STPKs. The size of the square is proportional to the number of that STPK's substrates, which is also given in parentheses. The round nodes show the number of substrates phosphorylated by one or multiple STPKs. The thickness of the edge corresponds to the number of substrates. All regulated phosphosites met the following criteria: > 2-fold change in GoF and/or < -2-fold change in LoF, Ascore > 19, and p-value <0.005.
(G) Phosphorylation of members of the proteasome/Clp protein degradation systems by different STPKs.
(H) Phosphorylation of RNA polymerase subunits by different STPKs.

Based on the number of the substrates per STPK, we defined differences in STPK-specific activity. We observed one group of STPKs that had low basal activity in WT strain in broth culture. Specifically, we noted that PknD, F, I, K and L showed less than 30 hypo-phosphorylated sites when inactivated (**Figure 3-2 and 3-3A**). However, these silent kinases can be activated as shown by the larger number of hyper-phosphorylated proteins in the GoF mutants. A group of kinases showed higher basal activity in WT strain in broth culture, represented by PknB, E, and H (**Figure 3-2 and 3-3A**). This kinase group could also be further activated, as demonstrated by additional substrates identified in the GoF mutants. We next tested if STPK induction in the GoF strains may lead to off-target phosphorylation (**Figure 3-4**). We found that the average magnitude and number of phosphorylation changes did not correlate with the abundance of the respective kinases. For example, the PknG GoF strain, which showed the highest degree of STPK induction, generated only 43 phosphorylation changes. These data supported the idea that GoF leads to predominantly physiologic phosphorylation events. Overall, using both LoF and GoF mutants identified distinct substrate sets for each STPK.

Kinases interact with each other more than with other protein families and can profoundly affect each other's activity by phosphorylation. It is possible that some of the observed changes with kinase perturbation may be indirect, which may be difficult to discriminate from direct effects. Phosphosites that are hyperphosphorylated in the LoF or hypophosphorylated in the GoF cannot be explained by direct phosphorylation and likely represent indirect phosphorylation events (**Figure 3-2, 3-3B, 3-3C**). All STPKs

showed some of these indirect effects, and ~8% of all altered phosphosites were such indirect sites (**Figure3-3C**). Indirect substrates may also be among the substrates that met our direct substrate criteria, but these are indistinguishable from the direct substrates in our data.

Dual specificity STPKs

Protein Tyr phosphorylation was thought to be absent in *Mtb* but, in the last ten years, has now been conclusively demonstrated (94). There are no annotated canonical bacterial Tyr kinases in *Mtb*, so the identity of the Tyr kinases remains an open question. Based on *in vitro* biochemical activity, atypical kinases and dual specificity kinases among the STPKs have previously been suggested as candidates. Our data further expanded the total pTyr sites in *Mtb* and allowed us to test in a cellular context whether STPKs are dual specificity. Most STPK-specific phosphorylation showed a preference of Ser/Thr over Tyr residues, but we observed many kinases regulate pTyr to some level. 231 pTyr sites, or 24% of all pTyr detected in our study were affected by STPK perturbation (**Figure 3-3E**). PknK and PknL had the largest number of regulated pTyr, suggesting dual specificity, while PknG, I, and J had the least, implying they behave strictly as Ser/Thr protein kinases. Altogether, these data postulate some *Mtb* STPKs evolved dual activity like the dual specificity kinases in eukaryotes, rather than mediated by specialized bacterial Tyr kinases.

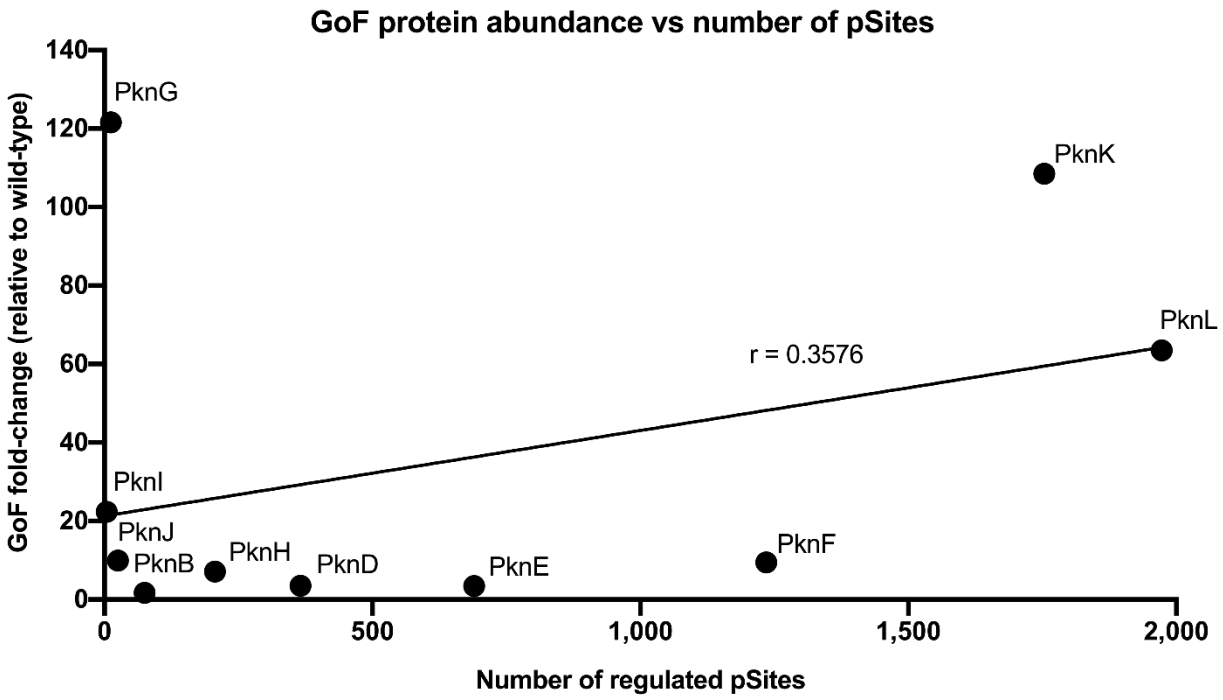


Figure 3- 4: Independence of STPK abundance and number of phosphorylation sites.

The abundance of STPKs in the GoF mutants does not overall correlate with the number of phosphosites detected in these strains, suggesting that GoF mutants retain substrate specificity and do not produce extensive off-target phosphorylation.

Phosphorylation Motifs

We next used the pLogo tool to identify potential substrate sequence preferences for each STPK (**Figure 3-3D, Figure 3-5**). We also identified sequence preferences for each phosphosite in the whole phosphoproteome. In both groups, we identified motifs for all phosphosites together (Ser, Thr, and Tyr) and residue-specific motifs. For all phosphosites in the whole phosphoproteome, there is a preference for negatively charged residues at the -1, -3, -4, -5, +1, +3 and positively charged residues are preferred in the +1 position. Here, we identified a general pattern of phosphorylated proteins.

There are clear differences in the STPK -specific motifs compared to the total phosphoproteome. For all STPKs, there are shared preferences: glycine in the +1 position, proline in the +2 position, and positively charged residues in the +5 position. For residues upstream of the phosphosite, we observed a dichotomy in STPK preference. PknD, PknF, and PknK, preferred negatively charged residues in the -1, -2, and -3 positions. However, PknE and PknL preferred positively charged residues at the -2 position.

Altogether, we predicted phosphorylation motifs around the phosphosites for PknE, PknF, PknK, and PknL. However, PknB, PknG, PknI, and PknJ, did not have a clear motif. For STPKs with a predicted motif, we next used a phosphorylation motif search program to determine if the motifs matched any known motifs (100). Based on this analysis, PknK and PknL vaguely matched the PKA kinase substrate motif

KXXX[pS/pT] (101). The PknF phosphorylation motif matched a casein kinase I substrate motif [E/D]XX[pS/pT] (102). Interestingly, PknE strongly matched to the Pim1 kinase substrate sequence [R/K][R/K][R/K]X[pS/pT]X, potentially identifying conservation between PknE and Pim1 kinase (103). Using these data, it is reasonable to screen Pim1 kinase inhibitors against PknE to establish this connection and potentially identify kinase inhibitors to target *Mtb*.

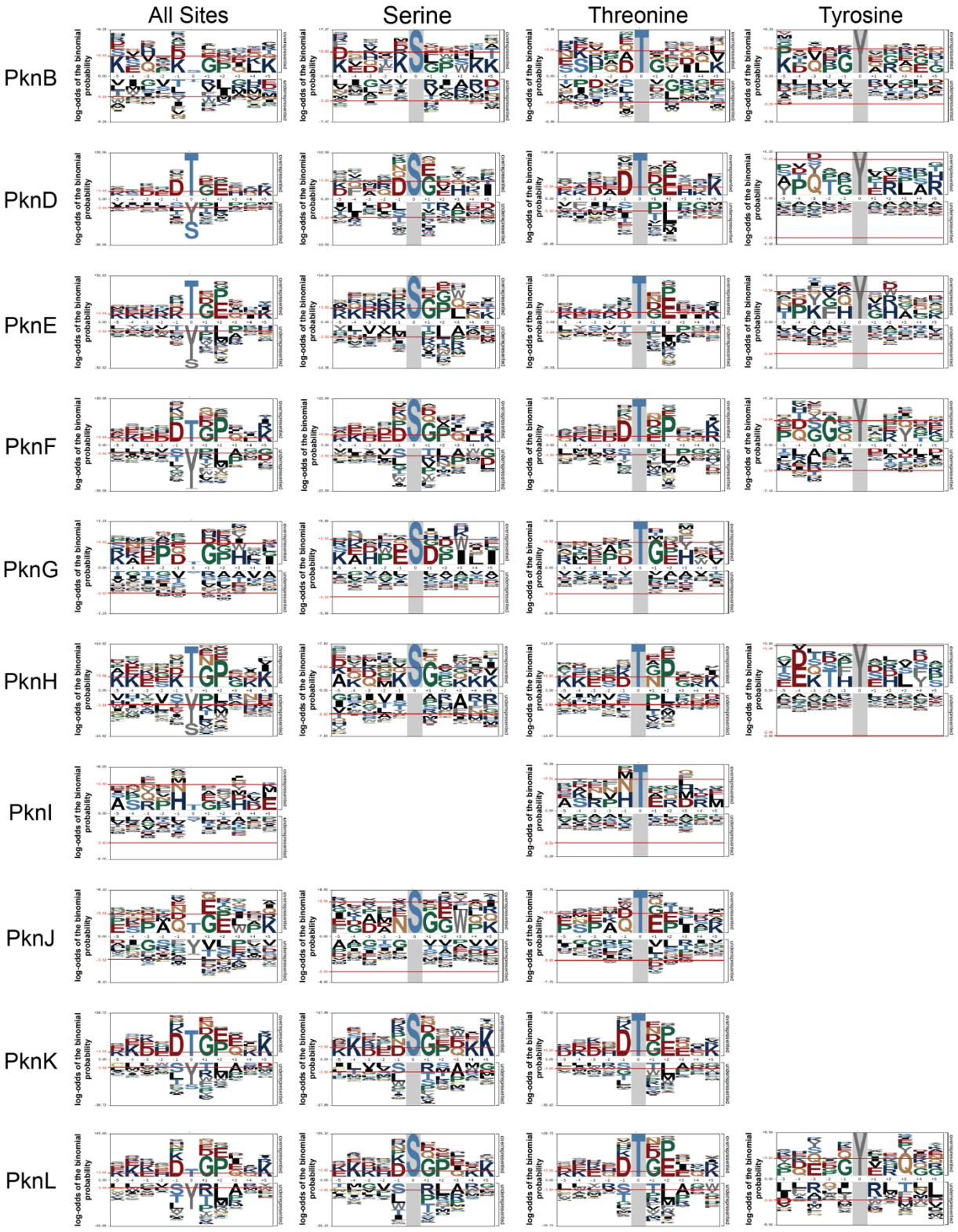

All Sites
Serine
Threonine
Tyrosine
PknB
PknD
PknE
PknF
PknG
PknH
PknI
PknJ
PknK
PknL
log-odds of the binomial probability

Figure 3- 5: STPK perturbation identifies STPK-specific phosphorylation motifs.

Substrates for each STPK were used to determine the preferred sequence motif. Rows show motifs for individual STPKs. Columns show motifs separated as all phosphosites, and then by Ser/Thr/Tyr-centered motifs. All regulated phosphosites met the following criteria: > 2-fold change in GoF and/or < -2-fold change in LoF, Ascore > 19, and p-value <0.005.

STPK – STPK phosphorylation

STPKs themselves are an important group of substrates to consider since STPKs are regulated by phosphorylation. We found that ten of the eleven STPK are substrates of another STPK (**Figure 3-6 and Figure 3-7**). The regulated phosphosites for all STPKs were distributed throughout the cytoplasmic domain. We identified six STPKs that contained regulated phosphosites in their kinase domains, the region containing the activation loop that regulates kinase activity (88, 104). Interestingly, for six STPKs, we detected regulated phosphosites in the juxta-membrane domain. Juxta-membrane phosphorylation sites are implicated in regulation or binding of accessory proteins, so these sites represent potential regulatory sites (105, 106). Four STPKs regulated phosphosites in the region N-terminal to the kinase domain, an area not well-studied for kinase activation or regulation. Altogether, these phosphosites are candidate STPK regulatory sites.

A previous study had explored interactions between STPK kinase domains *in vitro*, and in our study, we validated two of these interactions: PknB phosphorylates PknK and PknE phosphorylates PknD (88). Other interactions previous shown *in vitro* with recombinant protein did not replicate in our study. These discrepancies may be due to the use of kinase domain only in the *in vitro* studies or the lack of cellular context and illustrate the challenges to identify true cellular substrates *in vitro*.

We observed interesting patterns of phosphorylation between STPKs. PknG was not phosphorylated by or phosphorylated other STPKs. Both PknA and PknD are the most phosphorylated STPKs, and themselves substrates for six other STPKs each. PknF phosphorylated the most STPKs with six different substrate STPKs, although PknF did not have the most substrates overall. PknI only phosphorylated PknK and was not a substrate STPK. Even though PknI had the lowest number of substrates, it phosphorylated another STPK, as opposed to PknG that has more substrates and did not phosphorylate any other STPKs.

We also observed a few reciprocal STPK-STPK interactions: PknB-PknK, PknD-PknE, PknD-PknF, PknD-PknK, PknF-PknK, and PknK-PknL. PknD and PknK both had the greatest number of reciprocal interactions. Further work is needed to elucidate any regulatory role that these reciprocal interactions might play. Altogether, these data identified potential STPK regulation beyond the homodimerization and activation paradigm.

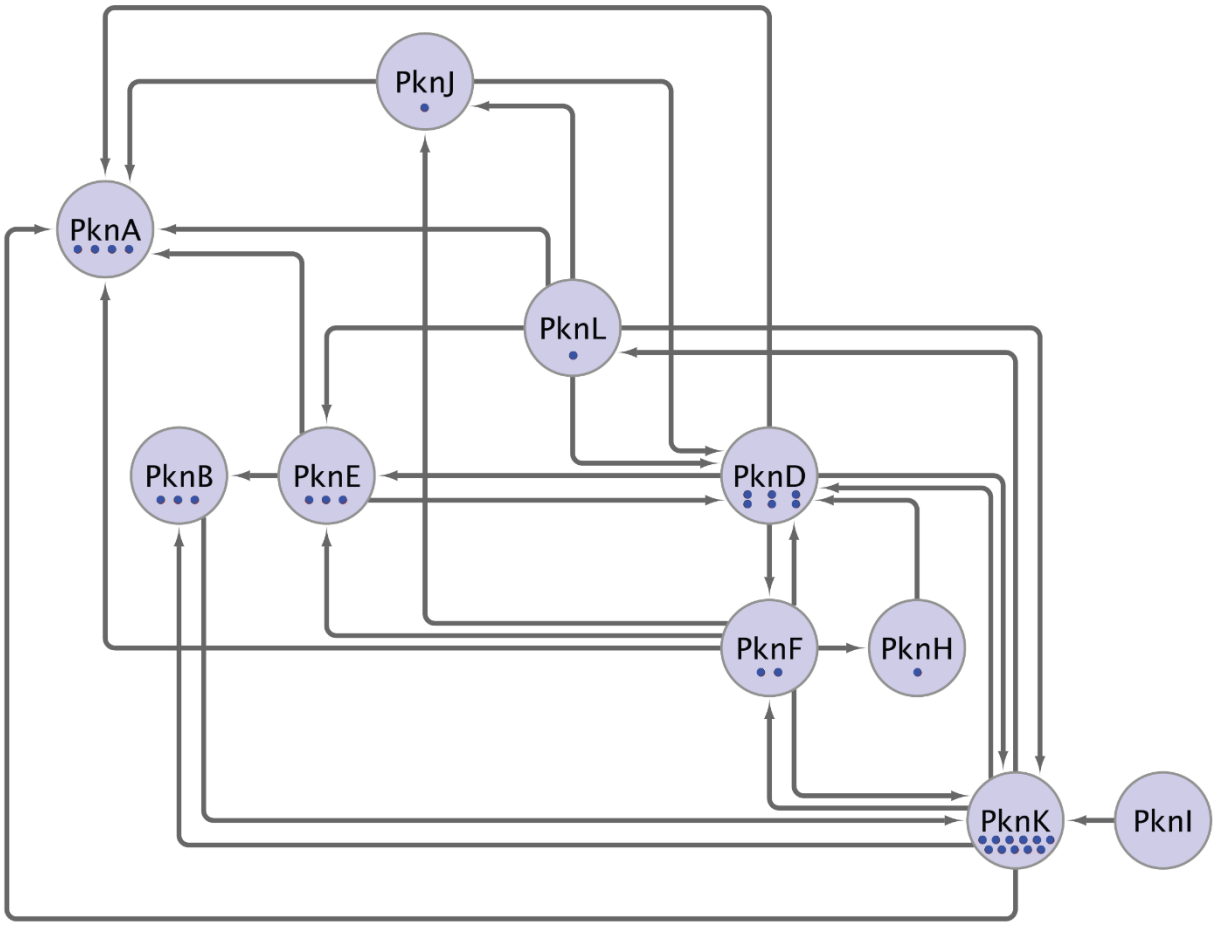


Figure 3- 6: STPK-STPK phosphorylation in *Mycobacterium tuberculosis*.

STPK - STPK interactions as shown using Cytoscape. Arrows indicate that the STPK at the base of the arrow phosphorylates the STPK at the end of the arrow. Blue dots denote the number of unique phosphosites in each STPK regulated by other STPKs. All interactions met a fold-change > 2-fold or < -2-fold, an ASCORE > 19, and a p-value < 0.005.

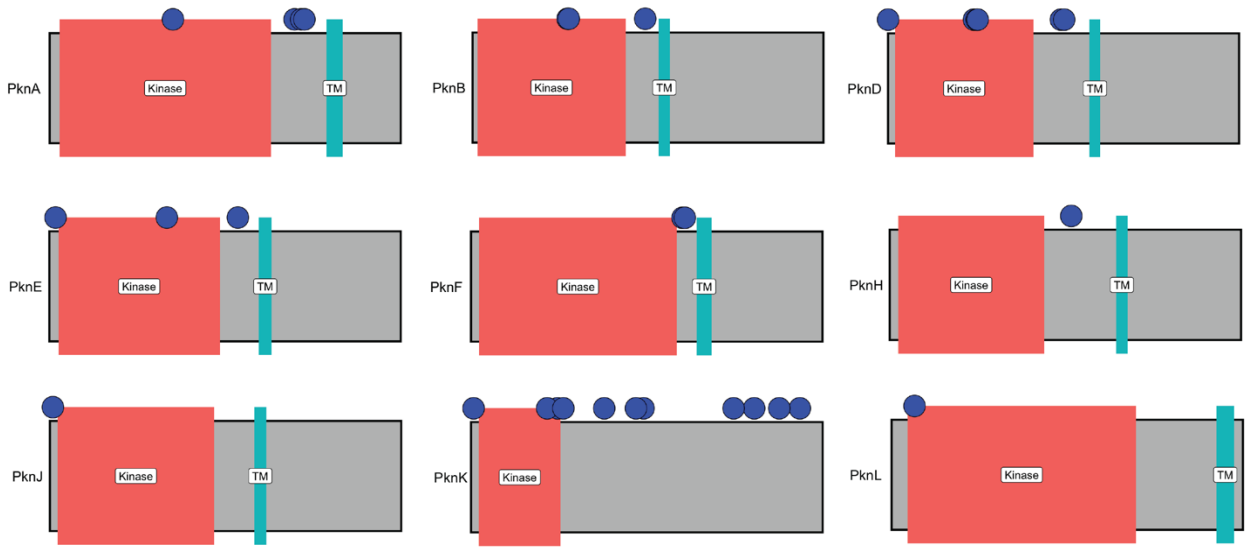


Figure 3- 7: Differentially phosphorylated phosphosites on STPKs.

Cartoons showing the domain organization of each STPK and differentially phosphorylated phosphosites, indicated by a blue circle. All phosphosites here met a fold-change > 2-fold or < -2-fold, an ASCORE > 19, and a p-value < 0.005.

The STPKs constitute a distributed network

The identification of multiple substrates per STPK and STPK-STPK interactions suggested a kinase signaling network. We next analyzed the kinase-substrate relationships globally to detect larger features of the network by visualizing all STPK-substrate interactions (**Figures 3-3F, 3-8**). We found that 469 substrate proteins were phosphorylated by a single STPK. Interestingly, we identified 935 proteins phosphorylated by more than one STPK. Further, we observed a minimum of two and up to ten STPKs phosphorylating a single substrate, with most substrates phosphorylated by two STPKs (293 proteins). Rv0020c was phosphorylated by ten STPKs, and, interestingly, has an FHA domain commonly shown to recognize phosphothreonine. In addition, Rv0020 also shows multi-site phosphorylation, implying Rv0020c could be a phospho-docking complex. We also analyzed the STPK substrates for enrichment in several protein families and KEGG terms. Among the most significantly enriched components are the proteasome and Clp protease systems, and RNA polymerase, almost all of which were phosphorylated by multiple STPKs, suggesting coordinated regulation by several inputs of these central cellular functions (**Figure 3-3 G and H**).

We also observed differences for each STPK in the proportion of substrates regulated solely by that STPK or by multiple STPKs (**Figure 3-8 and 3-9**). PknI only shared substrates with other STPKs. PknK and PknL solely phosphorylated more than 17% of their substrates. Altogether, these data showed that the STPK substrates are highly

redundant and crosstalk. The high redundancy may explain the paucity of *in vitro* phenotypes of KO strains in most STPKs despite the large number of substrates and is reminiscent of distributed eukaryotic phosphosignaling networks. Importantly, these overlapping substrate pools show that *Mtb* STPKs indeed form a kinase signaling network.

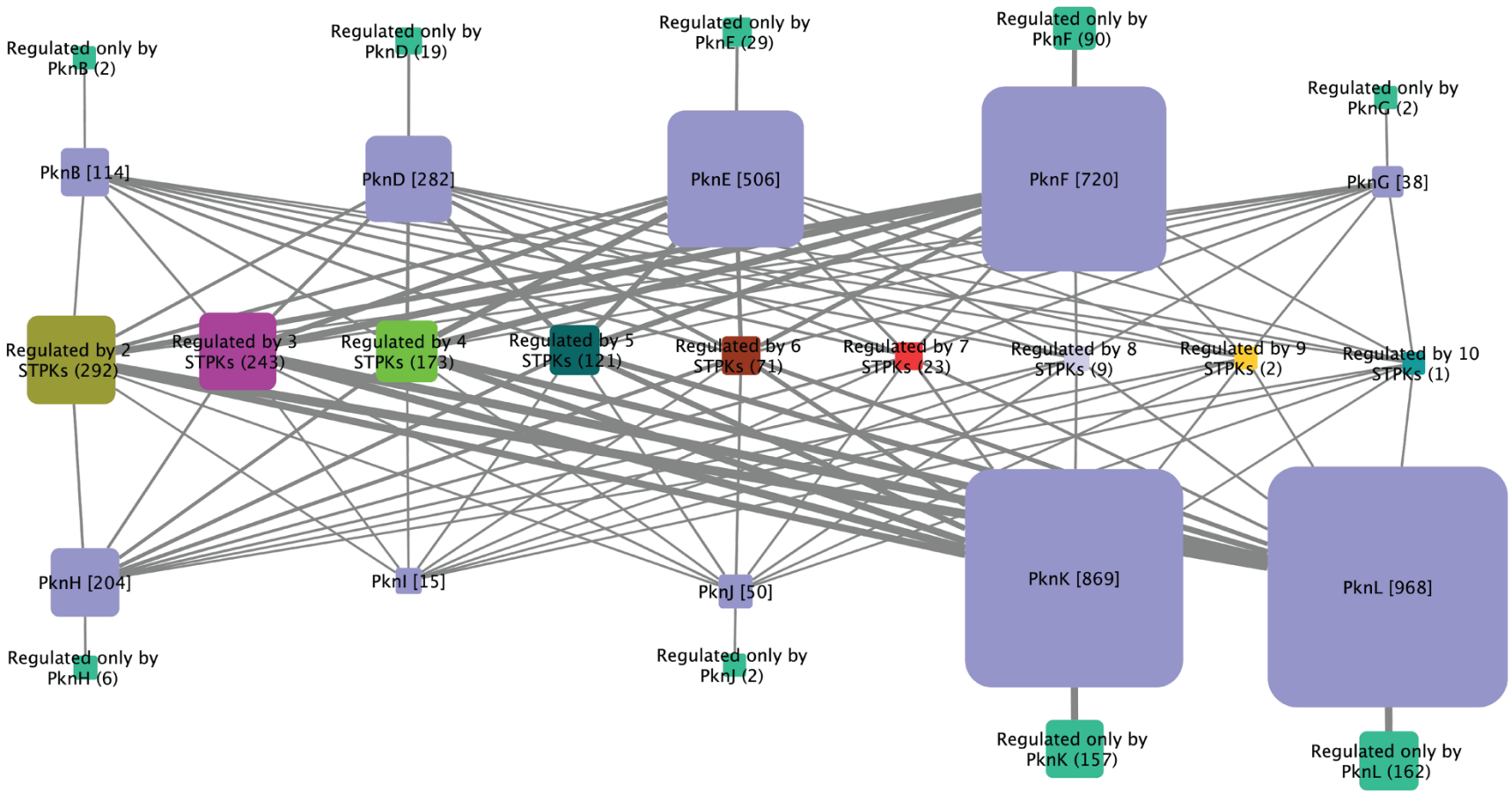

Regulated only by PknB (2)
Regulated only by PknD (19)
Regulated only by PknE (29)
Regulated only by PknF (90)
Regulated only by PknG (2)
PknB [114]
PknD [282]
PknE [506]
PknF [720]
PknG [38]
Regulated by 2 STPKs (292)
Regulated by 3 STPKs (243)
Regulated by 4 STPKs (173)
Regulated by 5 STPKs (121)
Regulated by 6 STPKs (71)
Regulated by 7 STPKs (23)
Regulated by 8 STPKs (9)
Regulated by 9 STPKs (2)
Regulated by 10 STPKs (1)
PknH [204]
PknI [15]
PknJ [50]
PknK [869]
PknL [968]
Regulated only by PknH (6)
Regulated only by PknJ (2)
Regulated only by PknK (157)
Regulated only by PknL (162)

Figure 3- 8: The *Mtb* O-phosphorylation is highly integrated.

An expanded global view of the STPKs' substrates and overlap in substrate phosphorylation. The purple squares represent the STPKs. The size of the square is proportional to the number of that STPK's substrates, which is also given in parentheses. The center squares represent substrates phosphorylated by multiple STPKs, differentiated by color and annotation. Lines connect STPK to each substrate pool and the width of the line indicates the contribution of each STPK to the specific pool. All regulated phosphosites met the following criteria: > 2-fold change in GoF and/or < -2-fold change in LoF, Ascore > 19, and p-value <0.005.

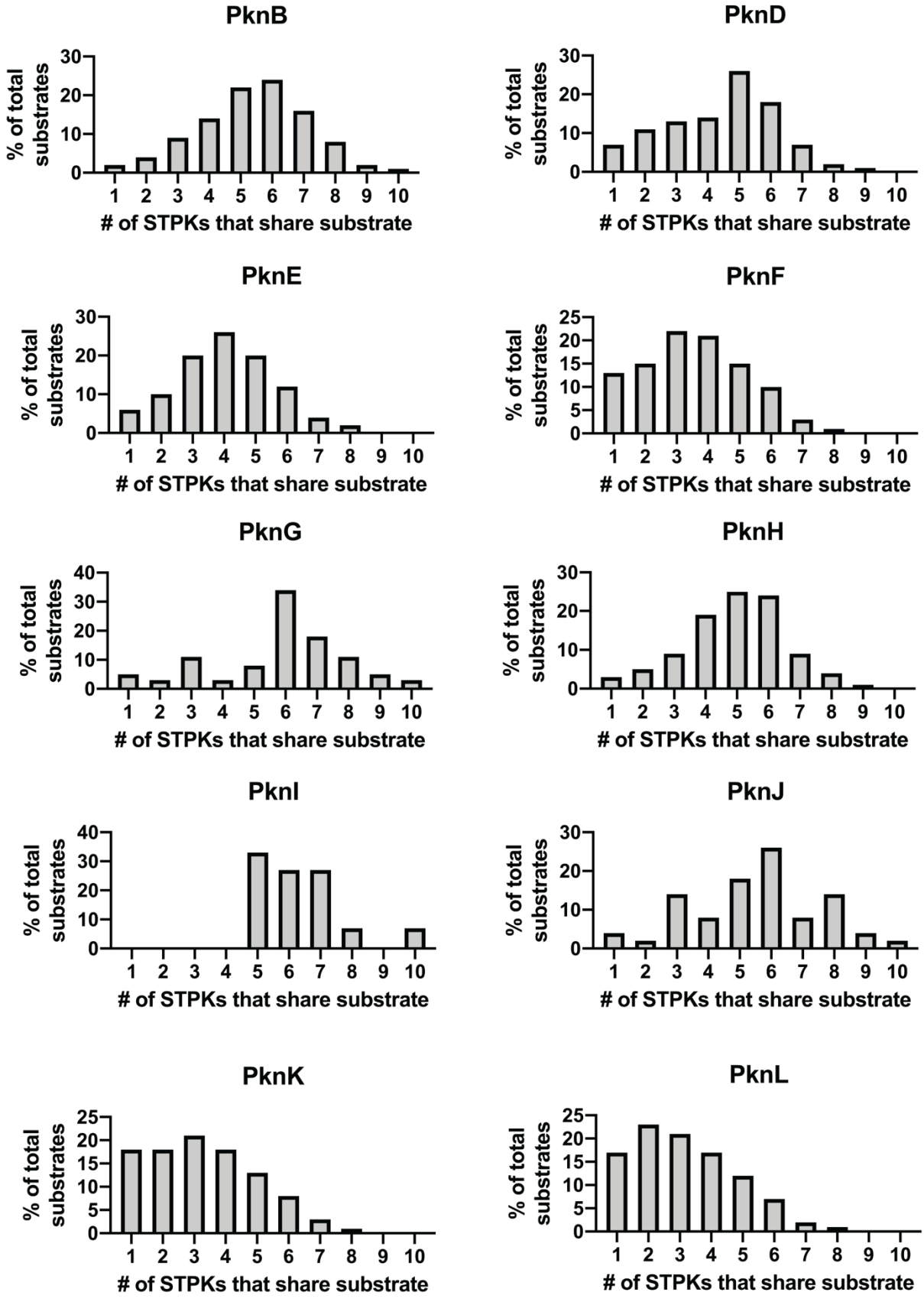

PknB
% of total substrates
# of STPKs that share substrate
PknD
% of total substrates
# of STPKs that share substrate
PknE
% of total substrates
# of STPKs that share substrate
PknF
% of total substrates
# of STPKs that share substrate
PknG
% of total substrates
# of STPKs that share substrate
PknH
% of total substrates
# of STPKs that share substrate
PknI
% of total substrates
# of STPKs that share substrate
PknJ
% of total substrates
# of STPKs that share substrate
PknK
% of total substrates
# of STPKs that share substrate
PknL
% of total substrates
# of STPKs that share substrate

Figure 3- 9: *Mtb O*-phosphorylation shows different degrees of interconnectedness.

Bar graphs showing the proportion of each STPK's substrates as phosphorylated by one or more other STPKs.

Pathway enrichment may reveal biological processes regulated by STPKs

We next performed Gene Ontology pathway enrichment analysis on each STPK substrate pool (**Figure 3-10**). We observed enrichment in a variety of biosynthetic pathways, including amino acid, fatty acid, and ribosome. In addition, we observed enrichment in metabolic pathways such as the tricarboxylic acid cycle and purine nucleotide metabolism. There was broad enrichment in processes involved in producing RNA and protein, such as transcription antitermination and translation. Interestingly, we detected enrichment for various environmental stresses that include heat, hypoxia, and oxidative stress. These pathways align with the function of STPKs in sensing the environment and may link specific STPKs to environmental stimuli. Lastly, we observed enrichment in protein secretion via type VII secretion, a pathway important for *Mtb* virulence (107). Overall, these pathways represent cellular functions regulated by *O*-phosphorylation.

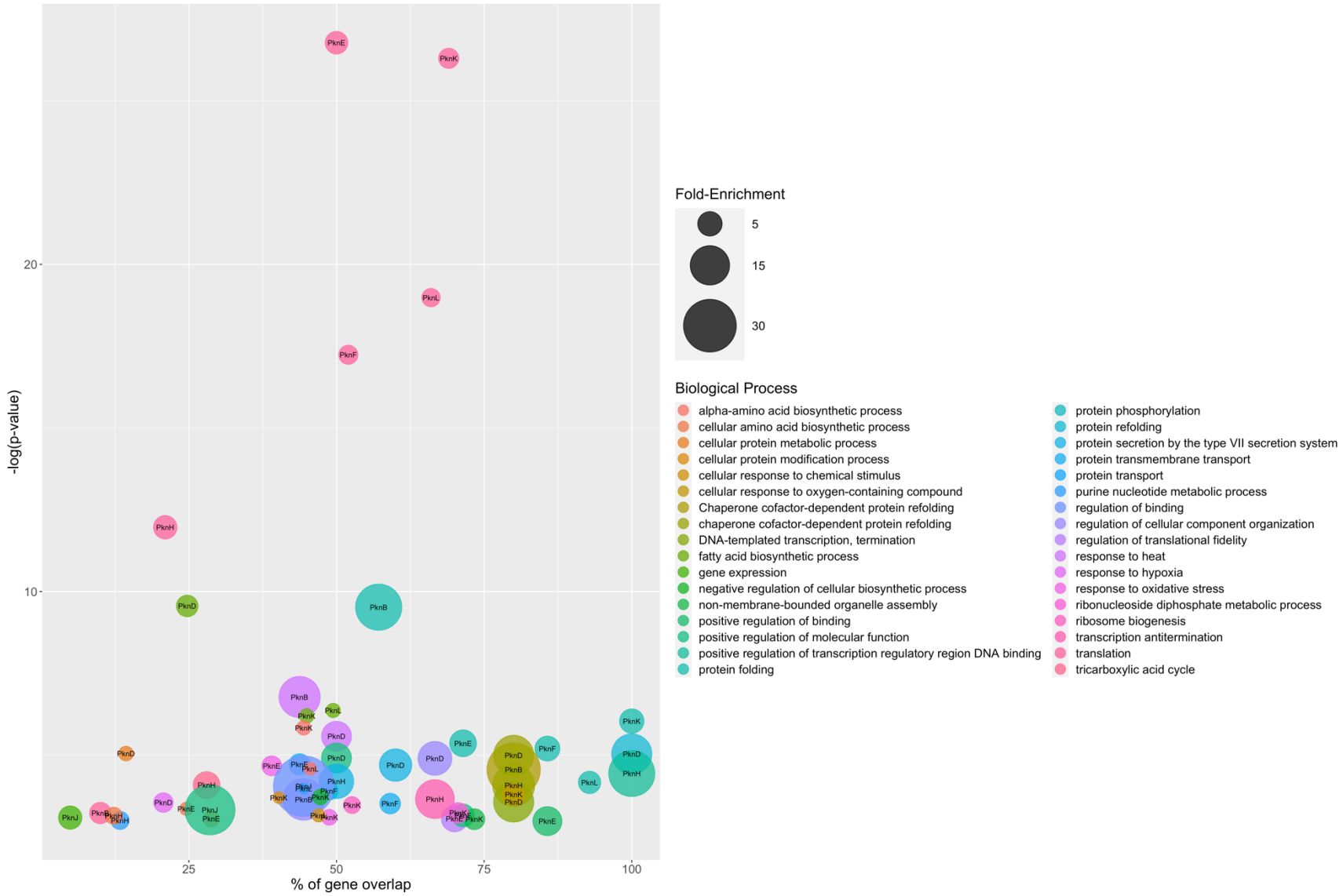

-log(p-value)
20
10
25
50
75
100
% of gene overlap
PknE
PknK
PknL
PknF
PknH
PknD
PknB
PknJ
Fold-Enrichment
5
15
30
Biological Process
alpha-amino acid biosynthetic process
cellular amino acid biosynthetic process
cellular protein metabolic process
cellular protein modification process
cellular response to chemical stimulus
cellular response to oxygen-containing compound
Chaperone cofactor-dependent protein refolding
chaperone cofactor-dependent protein refolding
DNA-templated transcription, termination
fatty acid biosynthetic process
gene expression
negative regulation of cellular biosynthetic process
non-membrane-bounded organelle assembly
positive regulation of binding
positive regulation of molecular function
positive regulation of transcription regulatory region DNA binding
protein folding
protein phosphorylation
protein refolding
protein secretion by the type VII secretion system
protein transmembrane transport
protein transport
purine nucleotide metabolic process
regulation of binding
regulation of cellular component organization
regulation of translational fidelity
response to heat
response to hypoxia
response to oxidative stress
ribonucleoside diphosphate metabolic process
ribosome biogenesis
transcription antitermination
translation
tricarboxylic acid cycle

Figure 3- 10: STPK substrates show enrichment in various *Mtb* pathways.

Gene Ontology enrichment analysis showing the enrichment of significantly regulated phosphoproteins for each STPK. The circle color denotes the biological process, the circle size indicates the fold-enrichment, and the annotation identifies the STPK enriched in a specific process. The x-axis shows the percent overlap of the STPK putative substrate pool and the annotated biological process. All regulated phosphosites met the following criteria: > 2-fold change in GoF and/or < -2-fold change in LoF, Ascore > 19, and p-value <0.005.

Discussion

TCSs are generally thought to be the major bacterial phosphosignaling systems, but much recent work has begun to focus on the importance of the STPKs. The study of STPK function has been severely limited by inadequate and technically challenging substrate identification methods that mostly relied on *in vitro* assays. As a result, the true size of the *Mtb* phosphoproteome remained an open question. Our work now showed that the *Mtb* phosphoproteome is at least five times larger than previously described. Our method of using both inactivating and activating mutants to identify STPK substrates identified over 12,000 new phosphosites. We showed that each kinase has different sized substrate pools during stationary phase. However, we also observed a pattern of redundant substrates and pathways between STPKs. The number of *O*-phosphorylated proteins approaches and even exceeds those in even the most highly phosphorylated eukaryotes (47, 108). No comparable kinase signaling network has been shown in bacteria, and this work may stimulate similar studies in other bacteria.

Our data recapitulated some previous STPK substrate identifications, but not others. We validated over 50% of reported interactions (58-60, 63, 65-67, 109-123), although some were below our p-value threshold. The discrepancies in substrate identification may be due to the different methodologies employed by our study compared to previous studies. The previous studies relied mostly on *in vitro* methods that may not always represent physiological interactions in the cellular context. Similarly, our study validated ~50% of interactions identified using STPK inactivation or inhibition in combination with

mass spectrometry. This overlap is not surprising considering variability between mass spectrometry experiments and the different growth conditions in each study (124).

The interactions among the STPKs themselves are important for understanding their regulation and signaling pathways. Previous *in vitro* studies showed that STPKs form into a hierarchical kinase signaling network (88). Our data also showed that all activation loop phosphosites tested in that study are phosphorylated. Of the ten different STPK-STPK interactions identified by Baer et al., we found two of these interactions, specifically that PknB phosphorylates PknK and PknE phosphorylates PknD. These data may differ due to the experimental design and STPK constructs as described above. We identified more phosphorylation that included phosphosites outside of the activation loop, such as N-terminus, regions in the kinase domain outside the activation loop, and juxta-membrane. These additional phosphosites may indicate novel mechanisms for STPK regulation, at least for *Mtb*.

Multi-phosphorylated proteins were common. Multisite phosphorylation can act as a docking mechanism to enhance strength and specificity for an interaction (125, 126). These sites can function in activation requiring multisite phosphorylation for maximal enzymatic activity (127). Structurally, multisite phosphorylation can result in different levels of protein stability (128, 129). Additionally, multisite phosphorylation can also function as a chelator, for example in the case of phosvitin (130). It is possible that *Mtb* multisite phosphorylation has similar functions.

Our work established STPKs as a major signaling system in *Mtb* that may be equal or larger than that of the TCSs. Further, our results build on several previous phosphoproteomic studies and expand the number of total phosphosites and STPK-substrate interactions. We provide a clearer understanding of STPK signaling in the cellular context than could be gleaned from *in vitro* experiments.

There are several limitations of our approach. First, our study identified substrates only during stationary phase. While the stationary phase has been shown to accumulate higher numbers of phosphorylated proteins in *E. coli* (91), it represents only one growth phase. Second, phosphorylation can be indirect. However, to some degree, we could parse these indirect from direct interactions as hypo-phosphorylated sites in GoF and hyper-phosphorylated site in LoF. Indirect sites that change in the same direction of direct sites cannot be distinguished by our data (or any other method), but we estimate that these are the minority based on the indirect sites we can measure (8% of total sites). Last, there are likely non-functional phosphosites. One study estimated that 65% of sites detected in phosphoproteomics experiments are non-functional (131). Computationally, functional phosphosites can be prioritized using residue conservation in substrate orthologs in related bacteria, but these predicted phosphosites still need to be experimentally characterized (132). Overall, this issue is hard to address without comprehensive functional assays or characterizing each phosphosite individually and poses perhaps the biggest current challenge to the field of posttranslational modifications in general.

CHAPTER 4: REGULATION OF *MTB* TRANSCRIPTION BY O-PHOSPHORYLATION

Introduction

In the previous chapter, we identified over 3,700 STPK-substrate interactions. These interactions link STPKs to a multitude of bacterial processes and are the foundation for dissecting the STPK signaling pathways. One prevalent type of STPK-substrate interaction was that of STPKs and transcription factors (TFs), suggesting a profound effect of STPK signaling on transcription. In this chapter, we broadly explore STPK regulation of transcription through TFs and the related TCSs.

Bacterial transcription is regulated on many levels. TFs in particular control RNA polymerase's interactions with target promoters, distributing RNA polymerase to different promoters in the genome and activating or inactivating transcription of the corresponding gene(s) (133, 134). TFs themselves are regulated by several mechanisms. Homodimerization and interactions of TFs with other DNA-binding proteins commonly affect their activity. Importantly, TFs can also be regulated by posttranslational modifications, such as acetylation and phosphorylation (135). Phosphorylation of TFs by STPKs or HKs affects DNA-binding, co-regulator binding, or the stability of the TF (136-139). Phosphoregulation of a TF's DNA-binding regions can create charge repulsion with the target DNA. In this way, STPKs can be directly linked to gene expression through the phosphorylation of TFs.

In *Mtb*, there are several known examples of TF regulation by phosphorylation, most of which involve the TCSs, in which a HK phosphorylates a RR in response to an extracellular signal. Examples of TF regulation by STPKs have also been reported, but only anecdotally. Several TF-STPK interactions including those between PknK and VirS, PknH and EmbR, and PknL and Rv2175c were identified only using *in vitro* phosphorylation assays (63, 65, 121). In another, more detailed example, PknB phosphorylated Lsr2, resulting in decreased DNA-binding and causing transcriptional changes important for growth in normoxic and hypoxic environments (67). Whether TF regulation by STPKs is a general mechanism of transcriptional regulation in *Mtb* is unclear.

Our phosphoproteomic data identified a large number of TF phosphosites that provided a unique opportunity to probe the intersection between STPK signaling and transcription. In addition, because we determined STPK-TF substrate pairs in combination with global functional readouts for the STPKs – transcriptional changes – our data offered an opportunity to predict the function of phosphorylation events on TFs. The identification of functional phosphosites among the thousands of sites now routinely identified in MS-based phosphoproteomic experiments such as ours is perhaps the greatest challenge in the phosphosignaling field today. As many as half of all phosphorylation sites are thought to be non-functional (131, 140). Identifying function of a phosphosite requires a priori knowledge of that function, individual phosphosite mutation, and function-specific assays, an approach that does not scale with the large numbers of phosphosites in current MS datasets.

The broad transcriptional effects of STPK signaling further provide functional context: The genes affected by STPK permutation provide immediate and testable hypotheses about the STPK's cellular functions. Accordingly, comparing any given gene set that may show altered transcription in a given experimental condition with that of an STPK permutation may identify the upstream STPK, thus connecting genes to upstream signaling events.

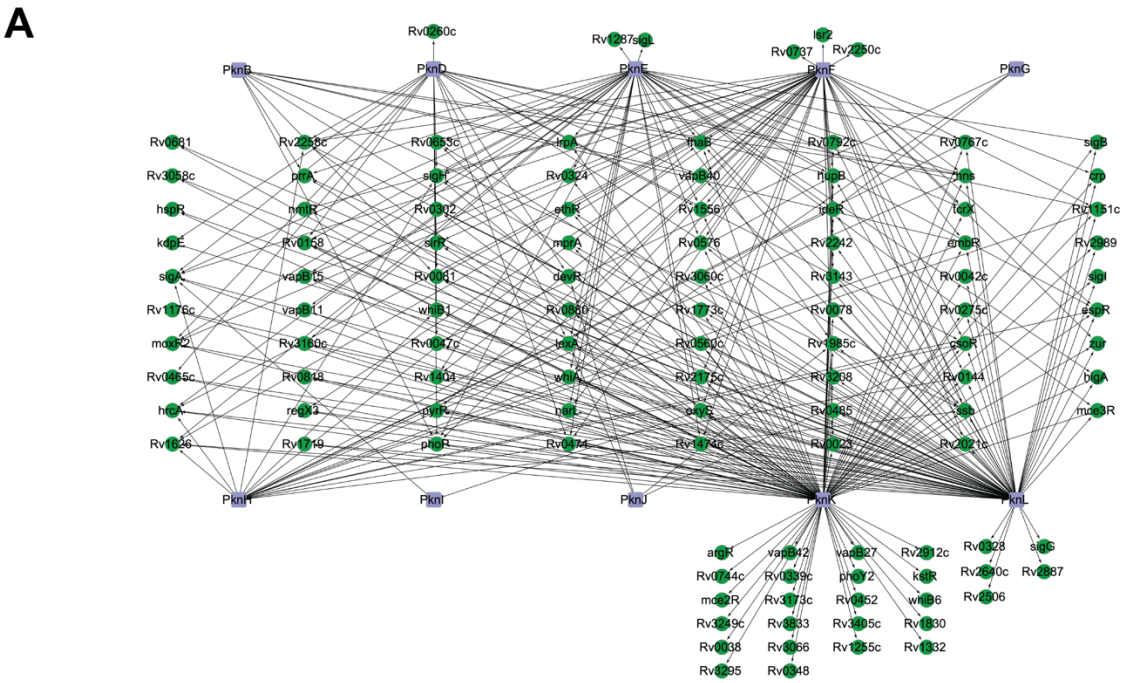
In this chapter, we use phosphoproteomic and transcriptomic analysis of STPK mutants described in chapter 3 to explore the role of *O*-phosphorylation in regulating transcription and showed that STPK perturbation results in large-scale changes in transcription. Further, integrating phosphoproteomic and transcriptional data, we predicted which of these STPK-TF interactions are functional. We validated one prediction by probing the interaction between PknK and Zur in detail. Interestingly, we also identified widespread STPK-TCS interactions and demonstrated that STPKs regulate HKs, placing STPKs above TCSs in the *Mtb* signaling hierarchy. Overall, our work demonstrated a large interface between STPKs and transcription in *Mtb* and that the STPKs and not the canonical TCSs are the main phosphosignaling enzymes controlling transcription.

Results

Transcription factors are STPK substrates

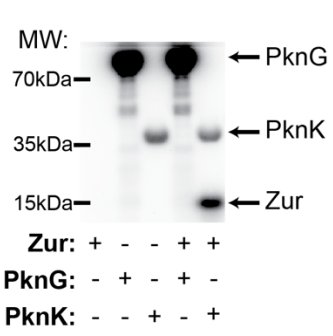
In our phosphoproteomic data, we found many different protein families represented. One group of highly phosphorylated proteins was the TF family. We identified *O*-phosphorylation on 158 TFs, representing 77% of TFs. We also found multiple phosphosites on a single TF, with as many as 20 sites on Rv1267c. Further, we identified 112 STPK-TF interactions, representing 50% of *Mtb* transcription factors (**Figure 4-1A, B**). There was a range in the number of TFs phosphorylated by an individual STPK with PknI phosphorylating 2 TFs and PknK 94. Altogether, these data showed that a large proportion of TFs are phosphorylated, and we linked over 70% of phosphorylated TFs to a specific STPK.

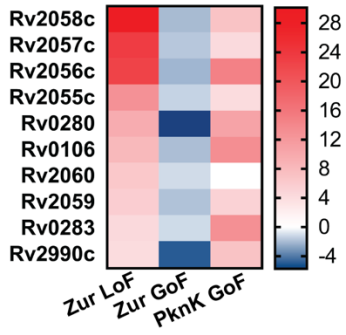
We next determined the location of all regulated phosphosites within TFs using TF structural predictions (**Figure 4-2**). Most phosphosites were in the DNA-binding domain, a common region for regulatory phosphorylation. Interestingly, we found phosphosites in protein-protein interaction or enzymatic domains, hinting at roles of phosphorylation beyond regulating TF-DNA binding. We observed 50% of the phosphosites in unstructured domains, including the N-terminal regions. Together, TFs were extensively phosphorylated in multiple regions, suggesting that regulation of *Mtb* TF activity through phosphorylation is a more prevalent regulatory mechanism than has been previously identified.

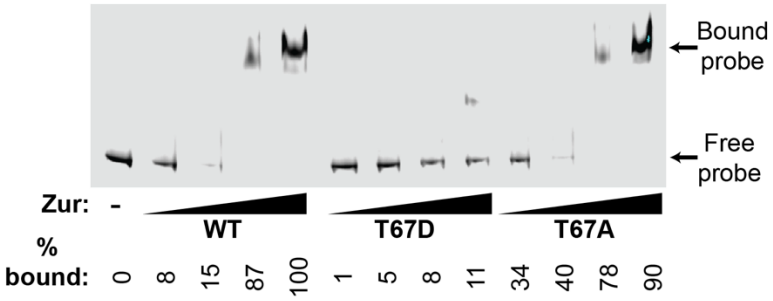


B

STPK	Number of TF substrates
PknB	9
PknD	21
PknE	39
PknF	48
PknG	3
PknH	20
PknI	2
PknJ	4
PknK	94
PknL	76







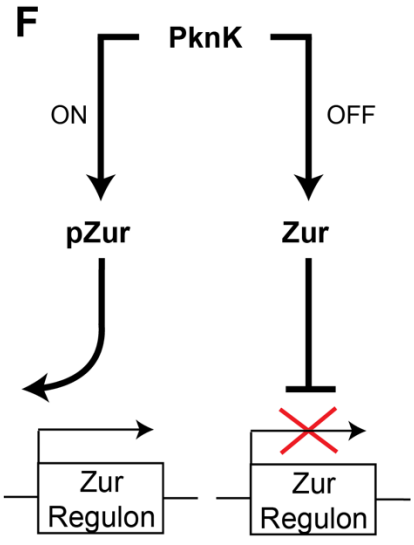


Figure 4- 1: PknK derepresses the transcription factor Zur.

(A) Overview of STPK-transcription factor interactions. STPKs are shown as purple squares, TFs are shown as green circles. Edges represent kinase-substrate interactions.
(B) Number of TF substrates of individual kinases.
(C) *In vitro* phosphorylation assay showing specific phosphorylation of recombinant Zur by PknK.
(D) Transcriptional response of Zur LoF, Zur GoF, and PknK GoF mutations on the Zur operon shows overlapping but opposite effects of PknK and Zur GoF, predicting inhibition of Zur transcriptional activity by PknK phosphorylation.
(E) Effect of Zur Thr67 phosphoablative and -mimetic mutations on DNA binding activity.
(F) Model of the PknK-Zur pathway.

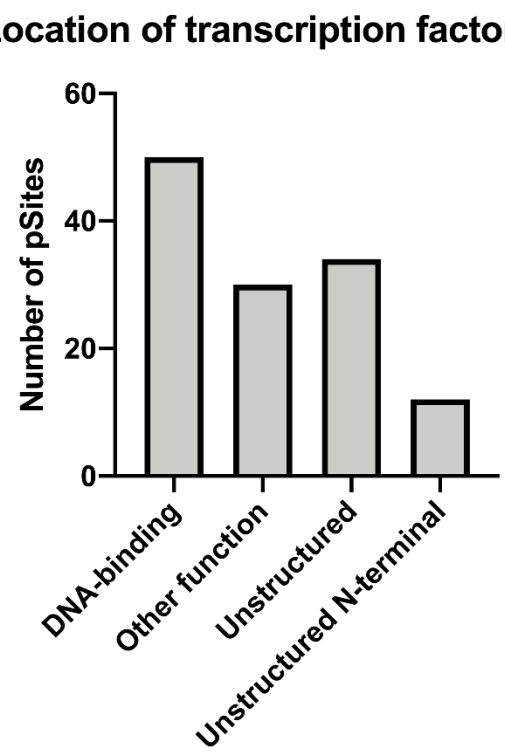


Figure 4- 2: Location of regulated phosphosites on TFs.

Bar graph showing the number of phosphosites that mapped to broad structural regions of transcription factors. All phosphosites used in this analysis met the following criteria: >2-fold change in GoF mutants and/or < -2-fold change in LoF mutants, ASCORE >19, and p-value < 0.005.

PknK regulates Zur via phosphorylation

To validate STPK-TF interactions, we explored the interaction between PknK and the zinc repressor Zur (Rv2359), a negative transcriptional regulator of metal homeostasis (**Figure 4-1C-F)**. As described in chapter 3, we measured the PknK mutant phosphoproteomes and compared them to WT. In a direct kinase-substrate interaction, we predicted substrates to be hypo-phosphorylated in the LoF mutant and/or hyper-phosphorylated in the GoF mutant. The PknK GoF mutant showed significant Zur hyperphosphorylation when compared to WT (8.7-fold, p-value=0.001), consistent with a direct kinase-substrate interaction. We validated this interaction using an *in vitro* kinase assay with recombinant PknK and Zur. Consistent with the phosphoproteomic data, recombinant PknK kinase domain efficiently phosphorylated Zur *in vitro*, but the negative control PknG, which did not affect Zur phosphorylation in our data, did not *(***Figure 4-1C***).*

We compared the previously identified Zur GoF regulon to genes regulated in the PknK GoF and found that >80% of the Zur GoF regulon are also regulated in the PknK GoF mutant (**Figure 4-1D**). The Zur GoF represses its regulon, confirming it as a repressor, but we observed the opposite with the PknK GoF mutant, which showed induction. Further, we compared PknK GoF gene expression changes to a previously defined Zur LoF mutant and found that both show Zur regulon induction. The Zur LoF data further supported Zur's role as a repressor, and our data show that PknK negatively regulates

Zur. Together, these data imply that Zur phosphorylation by PknK derepresses the Zur regulon and increased gene expression.

We next assessed whether Zur phosphorylation affects its ability to bind DNA. We made various recombinant Zur proteins with phosphosite mutations: WT, phosphomimetic (Thr67Asp), and phosphoablative (Thr67Ala). We next tested each protein's DNA-binding by electrophoretic mobility shift assays (**Figure 4-1E**). WT Zur bound to DNA, resulting in a shift of the DNA probe. Compared to WT, the phosphoablative mutant showed similar DNA-binding. The phosphomimetic mutant, however, showed decreased DNA-binding compared to WT. These data show an effect of phosphorylation on DNA-binding: Zur phosphorylation leads to decreased DNA-binding. Together, these data show a phosphoregulation circuit in which PknK phosphorylates Zur, resulting in decreased Zur DNA-binding and de-repression of the Zur regulon.

Transcriptional changes in STPK mutants

The effects of most *Mtb* STPKs on transcription are unknown. Based on the extensive phosphorylation of TFs we hypothesized that STPKs have large transcriptional effects. To gain a complete understanding of the STPKs' transcriptional effects, we analyzed all STPK mutant strains by RNA-seq. We grew the 20 STPK mutant strains in triplicate to stationary phase in 7H9 medium, extracted RNA, and then performed RNA-seq. Following RNA-seq, we used a cutoff of at least 4-fold up- or down-regulation compared

to WT and a significance cutoff of $p < 0.01$ to identify differentially expressed genes (DEGs).

Overall, 1,155 of 4,030 total genes met our criteria for DEGs, representing ~30% of *Mtb* genes (**Figure 4-3**). All STPK caused gene expression changes. Two distinct gene expression patterns were apparent. Several STPKs (PknB, G) showed the largest change in gene expression upon LoF, suggesting a constitutive housekeeping function. PknB, though only modestly downregulated in our strain, showed 30-fold more changes in the knockdown over the PknB GoF strain (**Figure 4-3**). Some STPKs (PknD, F, J, K, L) had little to no expression changes in the LoF mutant but showed large changes in the GoF mutant. For example, both PknF and PknJ did not show significant DEGs in the LoF mutant, but both the GoF mutants had over 100 DEGs. These STPKs appeared to be inactive and likely required an activation signal absent from *in vitro* culture. Importantly, the GoF mutants illustrate the advantage of testing both types of mutants to identify STPKs' downstream transcriptional effects.

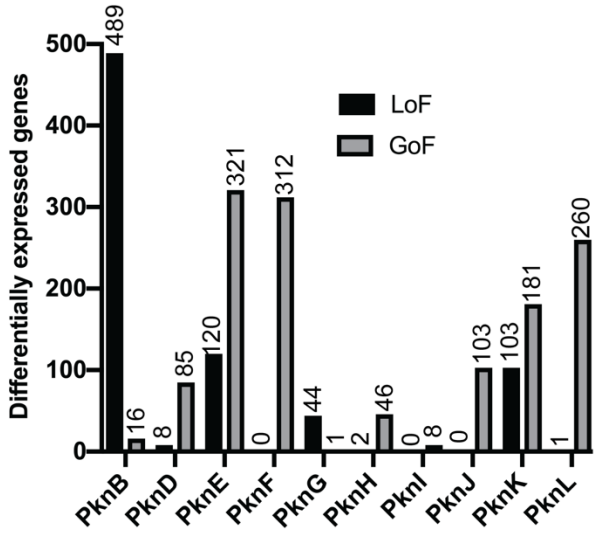
Among the LoF mutants, only PknB and PknG had more DEGs in the LoF than the GoF mutant, with the PknB LoF having the most with 489 DEGs (>10% of *Mtb* genes). All other STPKs showed more DEG in the GoF mutants. Interestingly, PknB and PknG both have growth phenotypes upon LoF, possibly reflecting the broad effect on transcription and multiple pathways. All GoF mutants showed changes in gene expression, indicating all were activated to some level by induction. All GoF strains, except PknB and PknG, had more DEGs in GoF than LoF, revealing that the GoF

mutants can identify transcriptional effects in these silent STPKs. The large changes upon GoF, however, raised the question whether induction leads to non-specific phosphorylation and artificial transcriptional effects. To test whether STPK expression correlates with total DEGs, we ranked the kinases by their degree of induction (**Figure 4-4A**). The number of differentially expressed genes was not correlated with the magnitude of STPK induction (r^2=0.37). To further explore potential non-specific effects of GoF mutations, we next asked if the abundance of STPKs in the GoF mutants were within physiologic levels. We mined existing transcriptional data for STPK expression during stress conditions and found that STPK expression levels in the GoF mutants were within the range found in WT when subjected to physiologic stresses such as hypoxia, oxidative stress, nitrosative stress, and exposure to arachidonic acid (**Figure 4-4B**). The expression levels of 8 STPKs in our GoF strains were in fact below those reported in WT under specific stresses, indicating that induction was not excessive, and the resulting phosphorylation and transcriptional effects are likely physiologic. Interestingly, the GoF expression data mirror our phosphoproteomic analysis showing some STPKs are inactive in culture and likely require a signal that is absent *in vitro*. Except for PknE and PknF, the number of differentially regulated phosphosites and the number of DEGs correlated well (r=0.92) (**Figure 4-4C**).

Integration of phosphoproteomic and transcriptional data predicts functional TF phosphosites

Predicting the share of functional phosphorylation events in any phosphoproteome has been a major challenge. Because we determined STPK-TF substrate pairs and global functional readouts for the STPKs by measuring transcriptional changes, our data offered an opportunity to predict function of phosphorylation events on TFs. We integrated our phosphoproteomic data on STPK-TF kinase-substrate pairs, the matched STPK transcriptional data, and a global transcriptional dataset mapping the transcriptional effects of ~200 *Mtb* TFs (141). We defined a functional interaction as a significant overlap between the regulon of an STPK and that of its cognate TF substrate. To correlate regulon of phosphorylated TFs and those of their cognate STPKs, we determined the genes altered by an STPK and its TF partner and plotted the % overlap of the TFs regulon with the STPK's regulon (**Figure 4-5**). 52 TFs showed significant regulon overlap (p=0.05) with their cognate STPK's regulon. 24 TFs showed > 50% of shared regulated genes, including PknK and Zur, and the regulons of 3 TFs (hns, pyrR, and Rv0792c) were fully contained in their STPK's regulon. These data predict a set of TFs whose phosphorylation is likely regulatory and leads to downstream gene expression changes.

A



B

Most induced genes

STPK	Mutant	Gene	Fold-change	p-value
PknH	GoF	mhpE	333	3.84e-43
PknH	GoF	Rv1271c	259	2.43e-55
PknK	GoF	echA13	227	1.28e-85
PknH	GoF	Rv1291c	168	1.89e-33
PknK	GoF	whiB5	164	4.56e-32
PknK	GoF	fadE17	163	2.29e-52
PknK	GoF	Rv1936	109	3.07e-142
PknH	GoF	Rv1290A	68	4.68e-20
PknK	GoF	Rv0077c	61	2.69e-51
PknB	LoF	papA1	59	6.30e-26

Most repressed genes

STPK	Mutant	Gene	Fold-change	p-value
PknL	GoF	nirB	-89	0.00e+00
PknL	GoF	Rv3371	-72	0.00e+00
PknB	LoF	Rv2625c	-58	1.58e-25
PknB	LoF	Rv2628	-55	5.52e-23
PknB	LoF	hrp1	-55	1.90e-13
PknB	LoF	TB31.7	-45	1.44e-17
PknB	LoF	Rv2624c	-38	1.68e-20
PknB	LoF	Rv3129	-38	4.08e-11
PknL	GoF	Rv3131	-33	8.74e-24
PknL	GoF	glbN	-32	0.00e+00

C

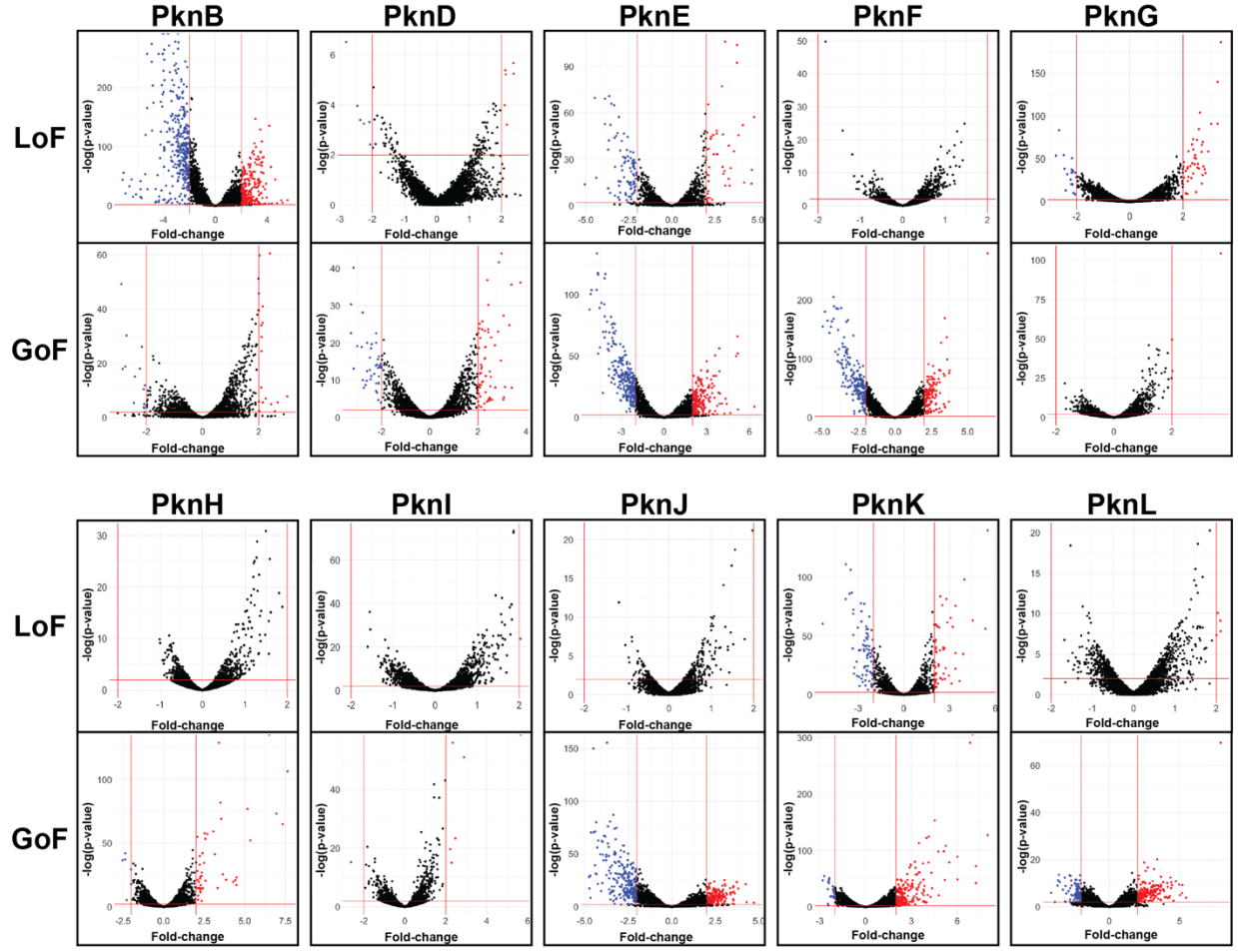


Figure 4- 3: *O*-phosphorylation has large transcriptional effects.

(A) Overall number of differentially expressed genes upon STPK perturbation by STPK and by type of perturbation. Differentially expressed genes are defined as >4-fold change with $p<0.005$.
(B) The most induced and repressed genes by STPK and type of perturbation.
(C) Types of transcriptional effects by STPK and perturbation. Volcano plot shows the DEGs in all STPK mutant strains, their direction, magnitude, and p-value associated with the change. Horizontal and vertical lines show significance cutoffs (>4-fold, $p<0.005$) used for further analysis of DEGs.

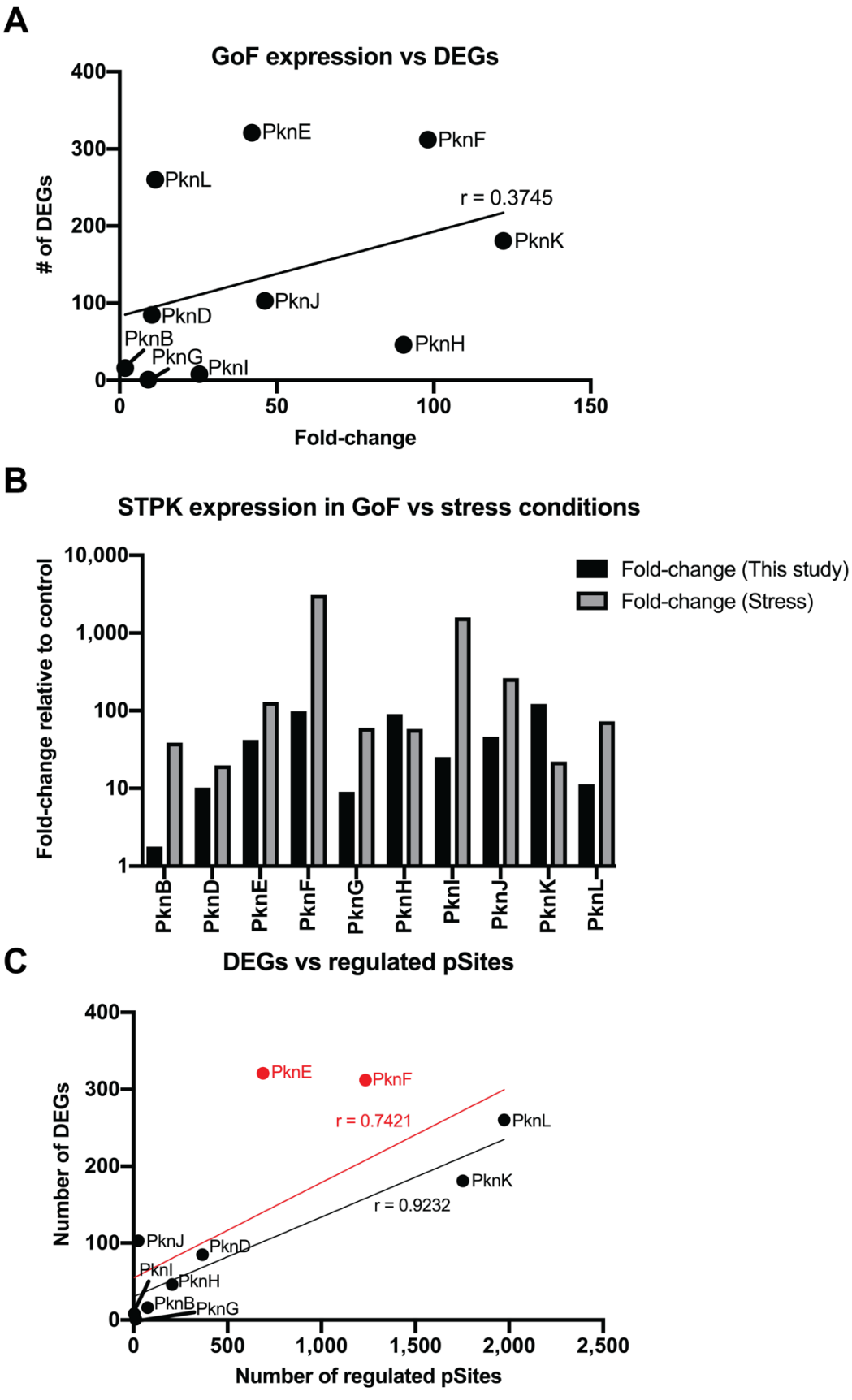

A
GoF expression vs DEGs
# of DEGs
Fold-change
PknE
PknF
PknL
PknK
PknJ
PknD
PknB
PknG
PknI
PknH
r = 0.3745
B
STPK expression in GoF vs stress conditions
Fold-change relative to control
Fold-change (This study)
Fold-change (Stress)
PknB
PknD
PknE
PknF
PknG
PknH
PknI
PknJ
PknK
PknL
C
DEGs vs regulated pSites
Number of DEGs
Number of regulated pSites
r = 0.7421
r = 0.9232

Figure 4- 4: Relationship between STPK induction, phosphosites, and DEGs.

(A) STPK abundance is not correlated with the number of DEGs in the GoF strains, indicating that STPK induction does not lead to global off-target changes in gene expression.
(B) The induction of STPKs in the GoF strains as determined by RNA-seq was within the range observed in historical data under physiologic stress conditions for most STPKs (142-146). PknK and PknL showed higher induction, most STPKs showed lower induction, suggesting physiologic levels of STPK.
(C) Except for PknE and PknF, more phosphorylation sites in the GoF mutant strains lead to more DEGs.

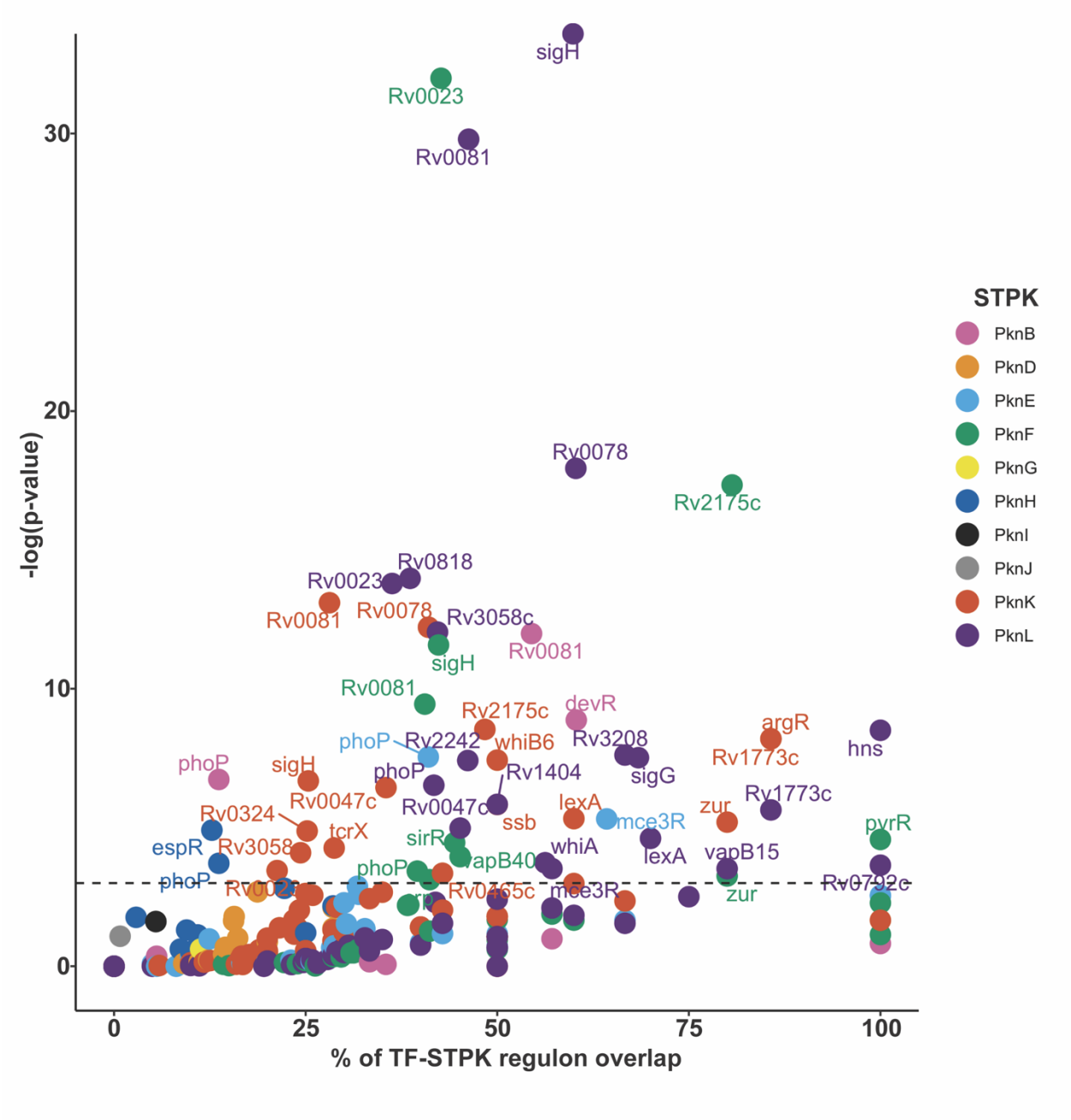


Figure 4- 5: Prediction of functional TF phosphorylation events.

Genes regulated by both TFs and their cognate STPK were plotted as % regulon overlap against the significance of the overlap. Colors denote different STPKs. Dotted line indicates a p-value of 0.005. These data are based on protein-level phosphorylation and no localization score cutoff was applied.

STPKs have large and overlapping transcriptional effects

To understand the higher-level organization of the STPKs' transcriptional effects, we visualized all DEGs (**Figure 4-6**). This visualization highlighted the large number of genes affected by STPKs and the high degree of transcriptional overlap between the STPKs. While there were 678 genes exclusively regulated by an individual STPK, 477 genes (41%) were affected by several STPKs. For example, transcription of Rv2011c was affected by three different LoF mutants, while Rv0975c was affected by six different GoF mutants. This gene co-regulation demonstrated that signaling by multiple *Mtb* STPKs converges on the same cellular pathways and suggested a high degree of redundancy. Further, we observed co-regulation in the same and in the opposite direction. For example, Rv2590 was induced in PknG and PknL LoF mutants while Rv1875 was repressed in PknE and PknK LoF mutants. In contrast, Rv1739c was induced in the PknE LoF mutant and repressed in the PknB LoF mutant. Interestingly, the connectedness of STPK gene expression given as the average number of STPKs that affect expression of one gene was on average 1.75. These data reveal a large, partially overlapping STPK signaling network with complex effects on gene expression.

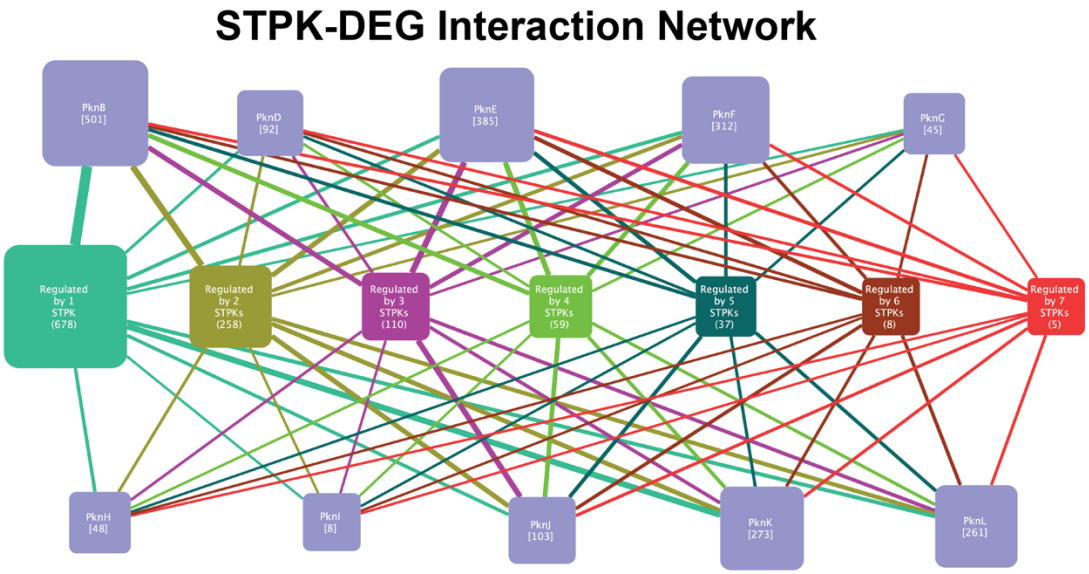


Figure 4- 6: STPK signaling is highly integrated with transcription.

STPKs (purple nodes) affect the expression of a large number of genes (number in parentheses). Nodes in the center represent the groups of genes that are regulated by a different number of STPKs (number in parentheses indicates number of genes in that group), illustrating the redundancy and/or cooperativity of the network.

The network organization was widely different between GoF and LoF mutants. One difference was the magnitude of repression and induction between STPKs and the genes they regulate. While both types of kinase perturbation could repress and activate gene expression, induction was more pronounced in the GoF than the LoF mutants (**Figure 4-3B**). However, the total share of repressed and induced genes was similar in the LoF (64% induced and 36% repressed) and GoF (68% induced and 32% repressed). GoF mutants affected a larger cumulative number of genes compared to LoF mutants with 873 genes in the GoF and 637 genes in the LoF mutants. Overlap of gene regulatory effects was quantitatively different between GoF and LoF networks. 85% of total genes are regulated in a single LoF mutant while 15% are co-regulated. In the GoF mutants, 66% of genes were regulated by one STPK while 34% were co-regulated. We further explored similarity of expression profiles between STPK mutant strains by hierarchical clustering of DEGs (**Figure 4-7**). The STPK LoF mutants formed two distinct branches with PknB separate from the other nine STPKs. Interestingly, PknB is the only essential STPKs in this group, and this separation may reflect its non-redundant functions. While also separated into two groups, the STPK GoF mutants clustered evenly between the two branches. The hierarchical clustering analysis highlighted groups of kinases that may function similarly and redundantly.

Additionally, we analyzed each LoF and GoF pair for differences in repression and induction (**Figure 4-8**). PknB and PknG LoF resulted in more gene induction than repression and GoF in more repression. PknE and PknK LoF showed more repression and GoF more induction. PknD and PknH GoF and LoF strains were similar, with

mutants showing more repression and more induction, respectively. These data indicate different inherent propensities of these kinases to repress or induce transcription.

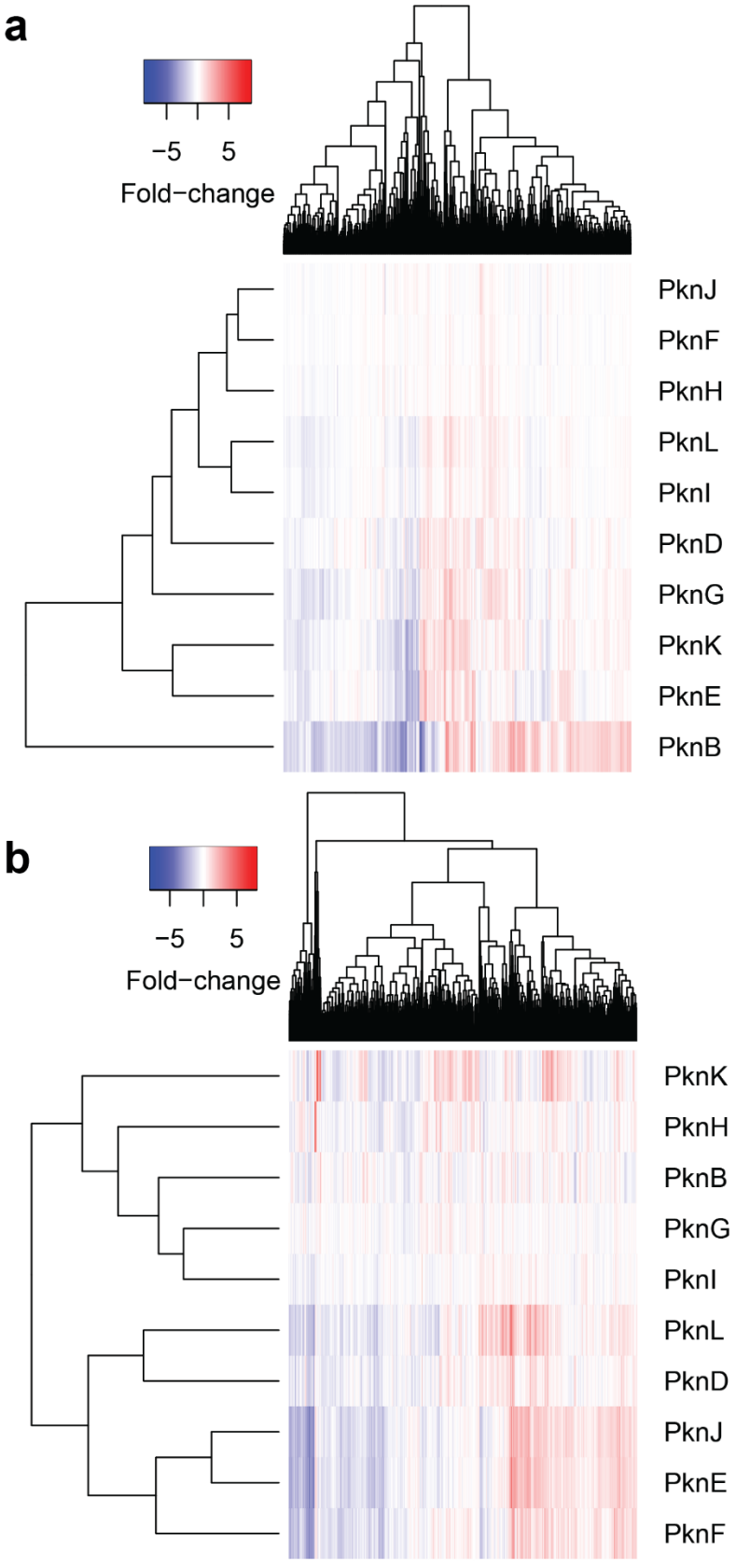

a
−5
5
Fold−change
PknJ
PknF
PknH
PknL
PknI
PknD
PknG
PknK
PknE
PknB
b
−5
5
Fold−change
PknK
PknH
PknB
PknG
PknI
PknL
PknD
PknJ
PknE
PknF

Figure 4- 7: STPK-mediated transcriptional changes reveal shared regulation between STPKs.

(A) Hierarchical clustering of gene expression changes in the STPK LoF mutants.
(B) Hierarchical clustering of gene expression changes in the STPK GoF mutants.

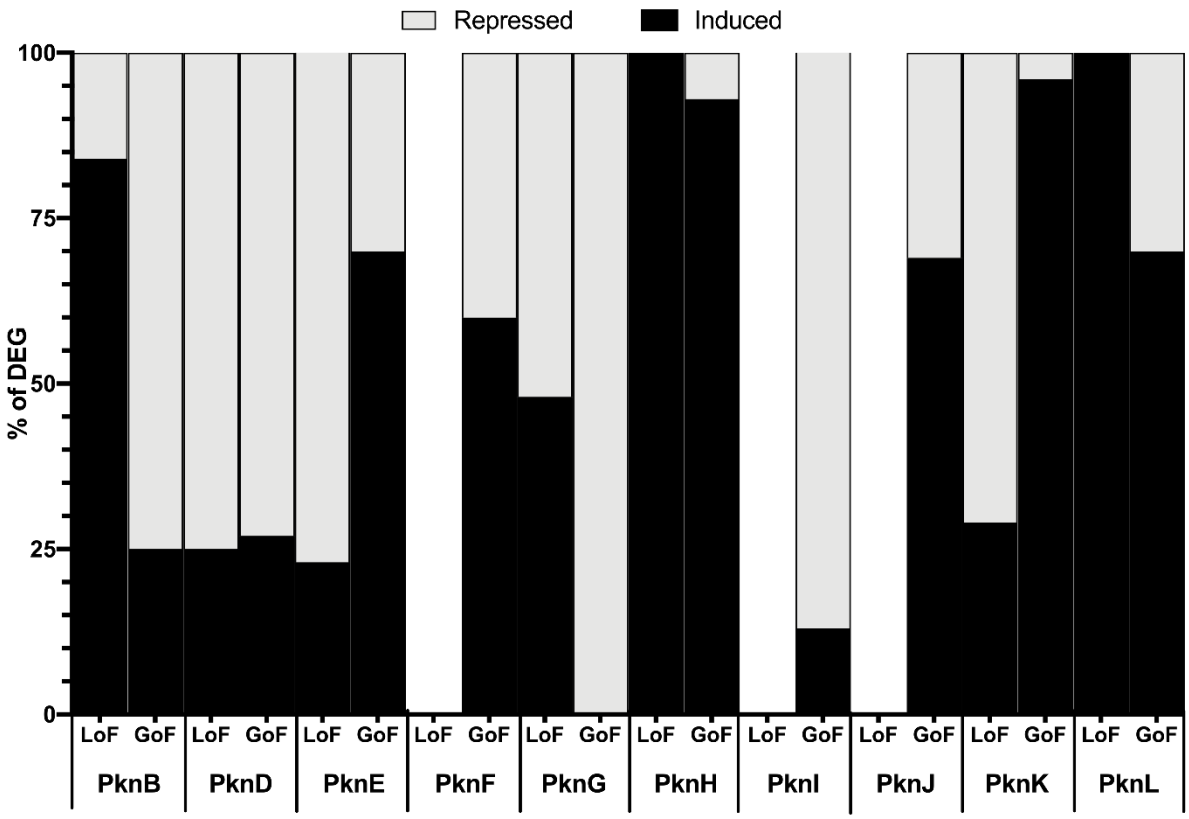


Figure 4- 8: STPKs show differential induction and repression between STPKs and mutants.

Bar graph showing the percentage of repressed and induced genes for each STPK mutant. Black bars indicate induction and gray bars indicate repression. No bars indicate a STPK mutants without significant effects on gene expression. All genes used in this analysis met all criteria: < -4-fold change or >4-fold change and p-value <0.01.

Pathway enrichment in differential gene expression

To broadly assign STPK effects to cellular pathways and to better predict kinase function, we used Gene Ontology (GO) enrichment analysis for each STPK mutant and their regulon (**Figure 4-9).** To this end, we plotted the p-value given by the GO analysis and calculated the specific % overlap between each STPK's regulon and the genes in that specific category. This representation identifies statistically significant enrichments as those towards the right and the top (**Fig. 4-9**). Although many STPKs were associated with GO pathways, six were significant (as defined by >50% overlap and p-value < 0.05) and potentially predict function.

Genes in the hydrocarbon catabolic processes category were enriched ~100-fold among the genes changing in PknD mutant strains. Genes in alkane catabolic processes and cholesterol catabolic processes were overrepresented in the regulons of PknE mutants and may link PknE to cholesterol utilization pathways required for survival in the host. Genes in copper ion transport were enriched in the regulons of PknG mutants. Interestingly, PknG phosphorylated the copper-inducible transcriptional regulator CsoR, suggesting a way in which PknG may regulate genes involved in copper transport. Genes in sulfate assimilation and arginine biosynthetic processes were enriched in PknK-regulated gene sets. PknK interacts with the arginine transcriptional repressor ArgR, consistent with the idea that PknK regulates arginine biosynthesis.

These enrichment analyses provide clues for the downstream pathways that may be regulated by each STPK, and when explored together with the phosphoproteomic data, can predict signaling pathways for individual STPKs.

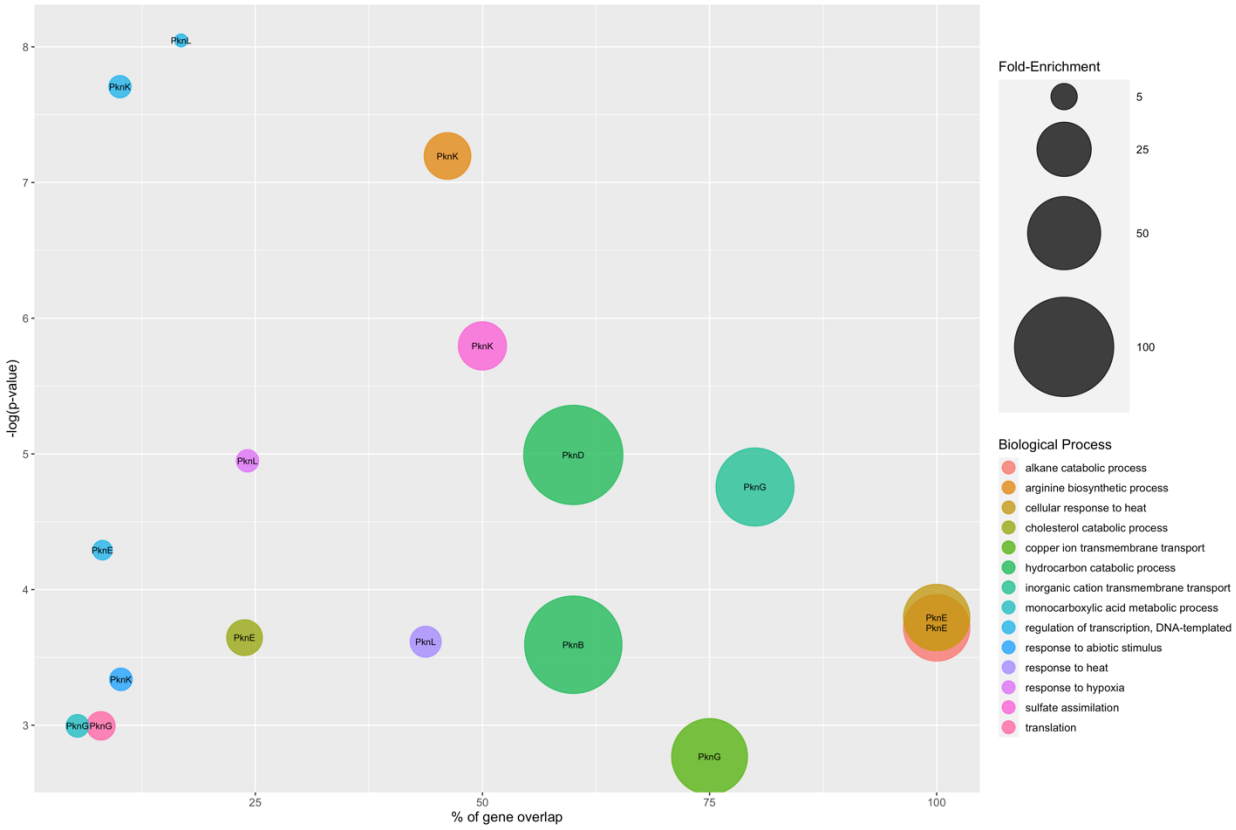


Figure 4- 9: DEGs in STPK perturbation reveals enrichment in *Mtb* pathways.

Gene Ontology Enrichment analysis showing the enrichment of significantly regulated DEGs for each STPK. The circle color denotes the biological process, the circle size indicates the fold-enrichment, and the annotation identifies the STPK enriched in a specific process. The x-axis shows the percent overlap of the STPK DEGs and the annotated biological process. All DEGs met the following criteria: > 4-fold change or < -4-fold change and p-value <0.01.

TCSs show extensive interactions with STPKs

There are currently only few known examples of an intersection between TCS and STPK signaling. In *Mtb*, crosstalk has only been reported between STPKs and RRs. Since we observed extensive transcriptional effects due to STPK perturbation and know TCSs regulate transcription, we next asked whether STPK and TCS signaling intersect by identifying STPK-TCS interactions in our data.

We observed 68 unique phosphosites on the HKs and 122 on the RRs (**Tables 2 and 3).** 11 of the 14 HKs (~79%) and 16 of the 17 RRs (~94%) were phosphorylated on Ser, Thr, or Tyr. For the HKs, we observed that a single protein can have 1-12 phosphosites, and for RRs, we observed 1-20 phosphosites. With kinase perturbation, we detected 49 unique instances of regulated phosphorylation, accounting for 6 of the 14 HKs (~43%) and 12 of the 17 RRs (~71%). In total, we noted that 9 of the 10 STPKs regulated at least one component of the TCSs (**Figure 4-10A**). These data described extensive interactions between STPKs and TCSs, including novel interactions between STPKs and HKs.

Rv Number	Gene Name	Num of pSites	pSites
Rv0490	senX3	3	S166, T258, S297
Rv0758	phoR	4	T193, S238, T255, T478
Rv0845	narS	5	S279, T286, T320, T361, T380
Rv0902c	prrB	6	T108, T276, S278, S285, T361, S94
Rv0982	mprB	12	T141, T205, S216, T245, S327, S387, T447, T482, S487, S490, S495, T504
Rv1028c	kdpD	5	T23, T24, S300, S553, S732
Rv1032c	trcS	1	T244
Rv2027c	dosT	6	S111, T2, S289, T331, T382, T555
Rv3132c	devS	11	S214, S23, T385, S450, S470, S483, S487, T512, T539, S541, S561
Rv3220c	Rv3220c	12	S103, S139, T144, T151, Y201, T297, T311, S335, S375, S455, S457, S468
Rv3245c	mtrB	3	S257, S304, S478

Table 2: HK phosphorylation.

The table shows all phosphosites identified on HKs. Red text indicates phosphosites regulated by STPKs.

Rv Number	Gene Name	Num of pSites	pSites
Rv0260c	Rv0260c	2	T342, T347
Rv0491	regX3	10	T151, T153, T174, Y187, T217, T79, Y98, T191, T100, T193
Rv0602c	tcrA	3	T107, T122, T48
Rv0757	phoP	20	T12, T9, T177, T112, Y118, T17, S173, S175, Y184, Y217, Y220, T235, T49, T51, T180, T191, S194, T120, Y241, T99
Rv0818	Rv0818	6	S123, S128, T138, S254, T151, T211
Rv0844c	narL	6	S135, S171, T185, S200, S156, S2
Rv0903c	prrA	10	Y108, S20, S33, T39, T6, S92, S93, T164, T225, S25
Rv0981	mprA	8	Y98, T141, T69, T79, T136, T119, S127, T215
Rv1027c	kdpE	9	T100, T120, T188, Y192, Y26, T30, Y196, T2, T72
Rv1033c	trcR	2	T206, T245
Rv1626	Rv1626	12	T5, Y111, S119, T144, S146, T169, T2, T6, S78, S95, S130, S82
Rv2884	Rv2884	1	S221
Rv3133c	devR	8	S123, S148, T198, S210, T180, S186, T205, Y184
Rv3143	Rv3143	5	T45, T70, T6, S5, S126
Rv3246c	mtrA	11	T213, T32, Y102, S132, Y185, T191, T217, T39, T43, S75, S97
Rv3765c	tcrX	9	T36, Y106, T108, S134, T37, S198, T222, Y228, T87

Table 3: RR phosphorylation.

The table shows all phosphosites identified on RRs. Red text indicates phosphosites regulated by STPKs.

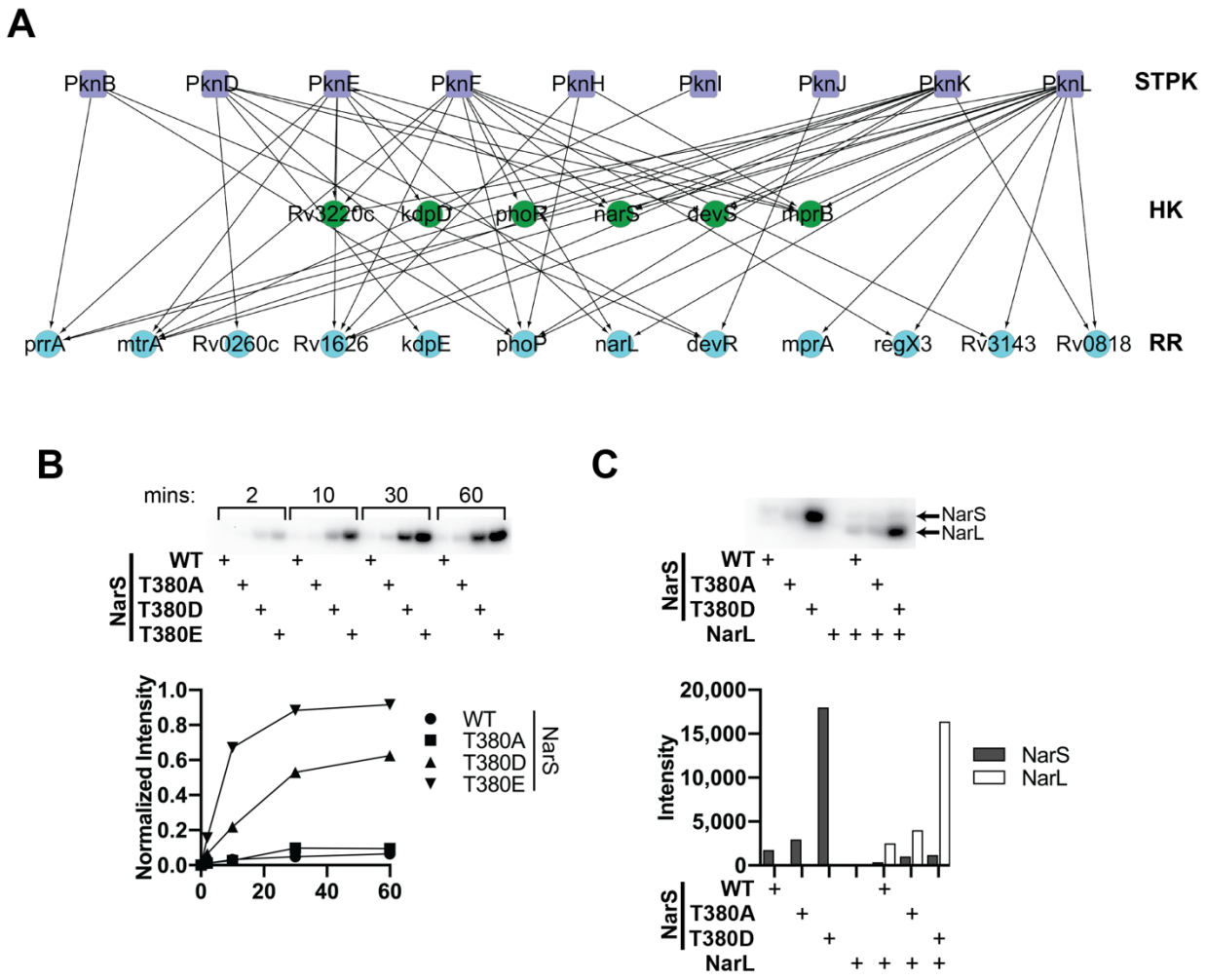


Figure 4- 10: The histidine kinase NarS is regulated by a STPK.

(A) Phosphorylation of TCSs by STPKs. Histidine kinases (green circles) and response regulators (blue circles) phosphorylated by the STPKs (purple squares) are shown.
(B) Phosphomimetic at Thr380 activate NarS. NarS autophosphorylation on His was visualized by [γ^{32}P]-ATP incorporation. The WT and phosphoablative Thr380Ala mutant show low auto-phosphorylating activity. The Thr380Asp and Thr380Glu phosphomimetic mutants show ~10-fold higher activity.
(C) NarS phosphomimetic (Thr380Asp) mutants have higher phosphotransfer activity to the cognate response regulator NarL than WT or a phosphoablative (Thr380Ala) mutant.

NarS is regulated by PknK

The HK NarS and RR NarL were differentially phosphorylated in our phosphoproteomics data. To test whether HKs are regulated by STPKs, we recombinantly expressed the soluble HK domain of NarS and its cognate RR NarL. (**Figure 4-10A**).

To test the effect of STPK phosphorylation on NarS activity, we generated Thr380 phosphomimetic (Thr380Asp and Thr380Glu) and phosphoablative mutants (Thr380Ala) and then measured autophosphorylation by [γ^{32}P]-ATP incorporation (**Figure 4-10B**). We found that WT and Thr380Ala showed similar levels of activity. Interestingly, both phosphomimetic mutants showed ~10-fold higher activity than both WT and Thr380Ala, demonstrating that phosphorylation by STPKs activates NarS.

We next tested whether the increase in auto-phosphorylation activity of the phosphomimetic mutants also resulted in functional phosphotransfer to the receiver aspartate of NarL (**Figure 4-10C**). We found that the Thr380Asp mutant was also efficient at transferring the phosphate from the HK to the RR, resulting in greater NarL phosphorylation. To our knowledge, this is the first example of activation of a HK by an STPK. These data suggest that some STPKs function upstream of the TCSs.

Discussion

The most direct connection between a signal received at the bacterial cell surface and a gene is through the phosphorylation of a TF. The TCSs are the canonical signaling pathway with this type of signal flow. TCSs consist of a receptor kinase with an extracellular sensor and intracellular kinase domain and often a single response regulator, which functions as a TF. Although this simple signaling pathway has long served as a paradigm of regulation of bacterial transcription, there is also an increasing number of examples of STPKs phosphorylating TFs, including in *Mtb*. Whether this type of STPK-regulated pathway is also common in *Mtb* is unknown, as are the relative contributions of STPKs and TCSs and their relationship to each other. In this study, we identified a surprisingly large number of *O*-phosphorylated TFs (50% of all TFs) that point to an extensive regulatory interface between STPKs and transcription. Of the STPK-TF interactions reported previously, we also detected those between PknH and DosR and PknL and Rv2175c. We observed that STPKs phosphorylated TFs on various domains. Classical TF phosphoregulation is due to phosphorylation of the DNA-binding domain, typically resulting in decreased DNA-binding. In our domain prediction, we indeed found many sites in DNA binding domains that are poised for such a negative regulatory effect. However, we also found phosphosites in regions outside the DNA-binding domain, implying other potential modes of regulation.

We analyzed the transcriptional effects of all but one *Mtb* STPK (PknA), testing two perturbations for each STPK. This near complete characterization of the STPKs' transcriptional effects provides a resource for connecting genes to STPKs: Any gene

expression change can now be linked to an upstream STPK. Together, these data provide new signaling context for the regulation of ~30% of the *Mtb* genome. A hallmark of STPK-regulated transcriptional effects is their redundancy. The STPKs have evolved into a coordinated system with largely overlapping effects, much unlike the TCSs. This type of signaling network can integrate several cues into one transcriptional response as well as one cue into several responses, providing convergent and divergent signaling effects. This type of highly connected network is typically associated with robustness, which is also reflected in the absence of *in vitro* phenotypes of most of the STPKs upon perturbation: Only PknA, B, and G show growth defects upon deletion, while the other 8 STPKs do not. Our phosphoproteomic and now also transcriptional data show that this lack of *in vitro* phenotypes despite large numbers of substrates and transcriptional effects is likely a result of the redundancy of the system.

There are a handful of examples showing that the *Mtb* STPKs regulate response regulators, such as PknH and DosR or PknB and RegX3 (147, 148), both interactions we also detected in our data. In addition, we detected many additional STPK-RR interactions in our dataset (35). We detected not only additional STPK-RR interactions, but also kinase-kinase interactions between these two systems that had been thought to be large insulated from and function independently of one another. Our data indicated that STPKs are upstream of HKs, and we did not identify STPK genes that were regulated by TCSs. The sweeping effects of the STPKs on gene expression, the numerous interactions between STPKs and TFs, and the control of HK activity by STPKs show a level of control of transcriptional regulation by STPK signaling that

exceeds that of the TCSs. These data challenge the notion of the TCSs as the main gene regulatory conduit and reposition STPK signaling to the center of *Mtb* gene regulation.

CHAPTER 5: CONCLUSIONs AND FUTURE DIRECTIONS

The studies discussed in this dissertation broadly explore the functions of STPKs in *Mtb*. We used MS-based phosphoproteomics in conjunction with kinase perturbation to identify global changes to the phosphoproteome. In Chapter 3, we identified substrates that changed in response to kinase inactivation or activation. In this first systems view of a bacterial STPK signaling network, we identified over 1,400 substrates and over >3,700 interactions regulated by STPKs which together formed a large, complex, interconnected, and distributed network previously only seen in eukaryotes. In Chapter 4, we hypothesized that *Mtb* STPKs regulate transcription. We found that STPK perturbation affected 30% of *Mtb* gene expression, and that these changes may be mediated by the direct phosphorylation of TFs. Surprisingly, we also observed many instances of TCSs being regulated by STPKs, the former representing the current paradigm for kinase's regulation of transcription. Overall, we provide a combined network of phosphoproteomic and transcriptomic changes due to kinase perturbation that illuminates the STPK system's architecture, predicts function, and provides insights into the understudied *Mtb* STPKs and bacterial STPKs in general (**Figure 5-1**).

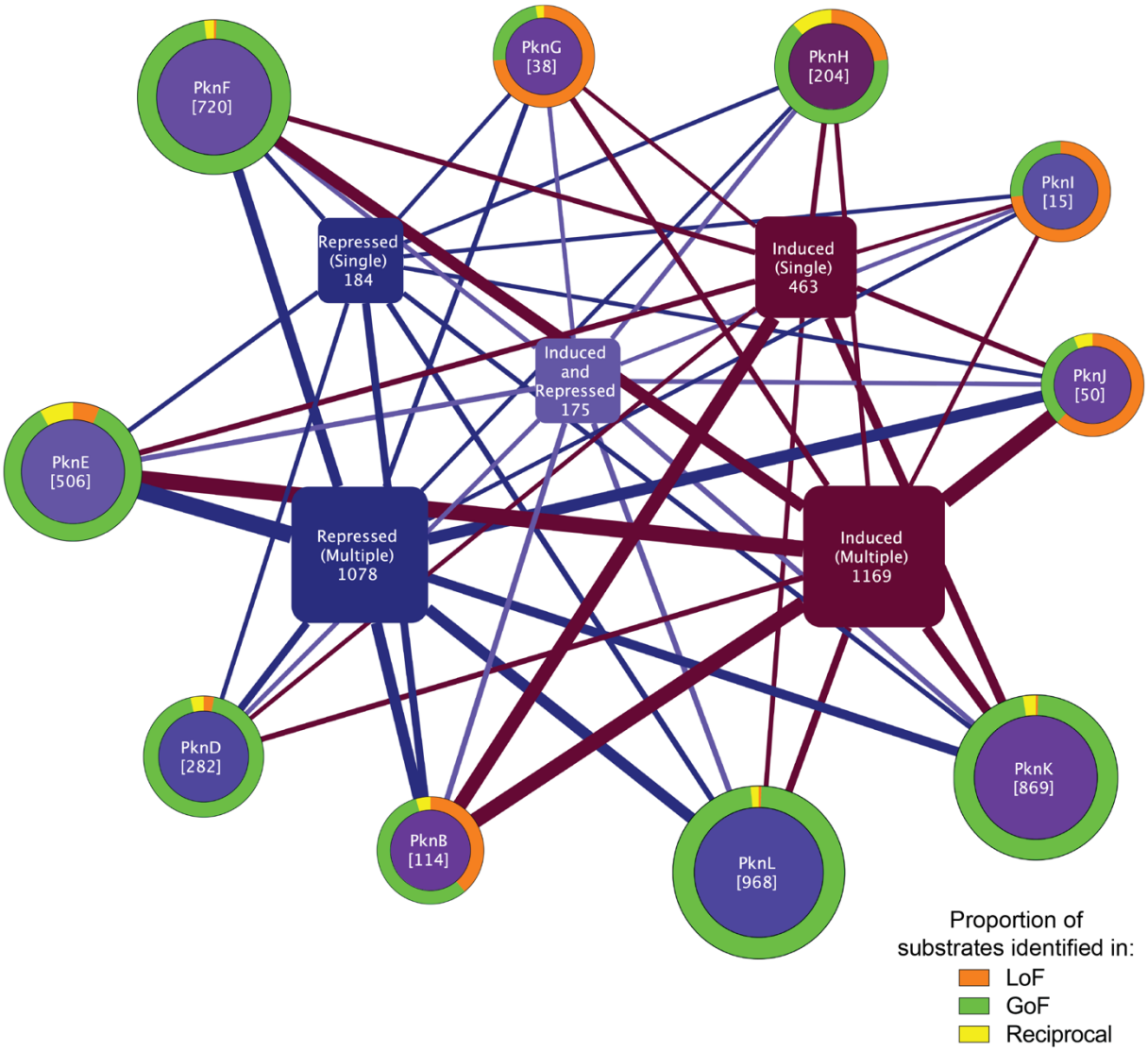


Figure 5- 1: Overview of the *Mtb* STPK *O*-phosphoproteome and transcriptional network.

The phosphoproteomic and transcriptional effects of STPK mutation are shown, illustrating the strongly overlapping nature of the network. The STPKs are shown as round nodes, with their total number of phosphorylated substrates in parentheses. The outer circle of each STPK node shows the relative share of substrates observed in the LoF (orange), GoF (green), or both (down in LoF, up in GoF, yellow). The inner square nodes represent genes regulated by the STPKs, with the thickness of edges connecting the STPKs to genes corresponding to the number of genes regulated by an STPK. Single and multiple refers to the number of STPKs that have the particular effect on the group of genes (for example, 1,078 genes are repressed by more than one STPK). Interestingly, the expression of 175 genes is activated by some STPKs and repressed by others.

Future Directions

For the past two decades, studies on *Mtb*'s STPKs focused on individual STPKs and often on *in vitro* assays for substrate identification. Our work built upon these previous studies by comprehensively analyzing 20 STPK LoF and GoF mutants within their cellular context to identify phosphorylation events and individual STPK-substrate interactions. We chose to probe phosphorylation in stationary phase because it provides a relatively steady state of the phosphoproteome and at least in *E. coli* is associated with the most phosphosites (91). In addition, slow or no growth in stationary phase is likely more representative *of in vivo* growth of *Mtb*. However, there are likely transient changes in the phosphoproteome associated with other growth phases that add another dynamic dimension to the phosphoproteome. Probing these transient changes and their potential functional consequences would complement the current data and likely discover additional functions. To obtain data for all STPKs in an activated GoF state, we chose to induce their expression. This strategy is in lieu of activation with the physiologic ligands, which are unknown in all but one case, or the respective stress that is sensed by the STPK, which is also unknown for most. It is likely that activation by the natural ligand or stress would most faithfully capture physiologic events downstream of the STPKs. To this end, our data provide a foundation that should inform the identification of STPK ligands. Many of the downstream events suggest upstream activators that can now be tested in targeted experiments. In the same way, the stresses sensed by the individual STPKs have come into clearer focus, although redundancy complicates the analysis of our data in this context.

This work will guide the identification of specific STPK functions. For virtually every aspect of *Mtb* biology, we identify at least parts of the putative upstream signaling context. We link STPK regulation to specific environmental conditions. For example, we show enrichment for PknD in genes related to hypoxia and PknE in genes related to oxidative stress, linking these STPKs to conditions encountered during infection. Further, we connect STPKs to host adaptation through phosphorylation of prrA, an RR associated with early adaptation to infection. In addition, these data also uncover novel targets for drug development: Not only the *in vitro* essential PknA and PknB might be tractable targets – other STPKs may be involved in drug efflux, drug resistance, or potentiate the action of existing drugs through regulation of proteins that activate drugs or are drug targets, such as katG (INH), rpoB (RIF), or rpsA (PZA).

This study revises our understanding of STPK signaling in *Mtb*. Do these findings extend to other bacteria? Given the large range of STPKs found in other bacteria from one to hundreds, the answer will depend on the bacterium. It is estimated that ~6% of all sequenced bacterial genomes have more STPKs than *Mtb*, which includes the pathogenic bacteria *Mycobacterium ulcerans* and *Nocardia farcinica (149)*. For those with a larger number of STPKs, our data suggest the presence of equally expansive, distributed, and cooperative signaling networks, the likes of which have so far only been described in eukaryotes.

Conclusions

Mtb remains one of the major causes of morbidity and mortality due to an infectious disease. Despite encountering many different host defenses, *Mtb* survives in the host by mechanisms still partly unknown. Protein kinase signaling is a key mechanism by which bacteria survive in the face of host defenses. The work detailed in this dissertation provides the molecular foundation for understanding a large swath of the signaling that underpins *Mtb's* adaptations to the host. This work provides an extensive inventory of the *O*-phosphorylation sites that collectively represent STPK signaling. We generated the deepest bacterial phosphoproteome to date by global MS-based phosphoproteomics. From >150 phosphoproteomes, we identified >5-fold more unique *Mtb O*-phosphorylation sites than previously reported. These data show that *O*-phosphorylation is in fact as prevalent in *Mtb* as it is in eukaryotes, with >70% of proteins phosphorylated. By generating a near complete STPK mutant panel of LoF and GoF strains of ten of the eleven Ser/Thr kinases followed by quantitative MS analysis, we assign >3,700 kinase-substrate interactions, providing a detailed map of the kinase-substrate relationships that determine each kinase's cellular function. These data represent the largest such database on STPK-substrate interactions to date. This work reports over 14,000 unique phosphosites and shows a prevalence of *O*-phosphorylation in *Mtb* that is on par with that of the most highly phosphorylated eukaryotic phosphoproteomes. In previous studies, the *Mtb* phosphoproteome has only been sampled superficially, and the striking prevalence of *O*-phosphorylation we identified suggests a much larger role in *Mtb* and possibly in bacteria in general as previously thought. The substrate-kinase interactions expand the number of known STPK-

substrate interactions from dozens to over 3,700 and provide a directory to assign STPKs to substrates and substrates to STPKs. The matched transcriptional data further assign transcriptional changes to individual kinases, connecting kinases to potential cellular functions. Together, this directory is a detailed map to assemble and understand the *Mtb O*-phosphorylation pathways that affect up to 70% of *Mtb* proteins and expression of ~30% of genes.

Beyond this substantial resource for *O*-phosphorylation and the STPKs' genetic effects, this study also provides new biological insight into this family of kinases, their signaling pathway structure, the overall network architecture, and the kinases' cellular functions: The control of ~30% of *Mtb* genes by STPK signaling makes them the likely dominant conduit for transcriptional regulation in *Mtb*. We show that direct phosphorylation of transcription factors by STPKs is a common signaling pathway structure. Not only do the STPKs control a larger swath of gene expression than the two component systems, but they are also likely to regulate the TCSs directly. These studies advance the understanding of STPK signaling in *Mtb* and will aid in understanding how it integrates with many different aspects of *Mtb* physiology and pathogenesis.

REFERENCES

1. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, et al. Tuberculosis. Nat Rev Dis Primers. 2016;2:16076.
2. Tiemersma EW, van der Werf MJ, Borgdorff MW, Williams BG, Nagelkerke NJ. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. PLoS One. 2011;6(4):e17601.
3. Global Tuberculosis Programme., World Health Organization. Global tuberculosis report 2021. Geneva, Switzerland: World Health Organisation; 2021.
4. Brennan PJ, Crick DC. The cell-wall core of Mycobacterium tuberculosis in the context of drug discovery. Curr Top Med Chem. 2007;7(5):475-88.
5. Maitra A, Munshi T, Healy J, Martin LT, Vollmer W, Keep NH, et al. Cell wall peptidoglycan in Mycobacterium tuberculosis: An Achilles' heel for the TB-causing pathogen. FEMS Microbiol Rev. 2019;43(5):548-75.
6. Romagnoli A, Etna MP, Giacomini E, Pardini M, Remoli ME, Corazzari M, et al. ESX-1 dependent impairment of autophagic flux by Mycobacterium tuberculosis in human dendritic cells. Autophagy. 2012;8(9):1357-70.
7. Simeone R, Bobard A, Lippmann J, Bitter W, Majlessi L, Brosch R, et al. Phagosomal rupture by Mycobacterium tuberculosis results in toxicity and host cell death. PLoS Pathog. 2012;8(2):e1002507.
8. Abdallah AM, Gey van Pittius NC, Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CM, et al. Type VII secretion--mycobacteria show the way. Nat Rev Microbiol. 2007;5(11):883-91.
9. Chan J, Xing Y, Magliozzo RS, Bloom BR. Killing of virulent Mycobacterium tuberculosis by reactive nitrogen intermediates produced by activated murine macrophages. J Exp Med. 1992;175(4):1111-22.
10. Flesch IE, Kaufmann SH. Activation of tuberculostatic macrophage functions by gamma interferon, interleukin-4, and tumor necrosis factor. Infect Immun. 1990;58(8):2675-7.
11. Stallings CL, Glickman MS. Is Mycobacterium tuberculosis stressed out? A critical assessment of the genetic evidence. Microbes Infect. 2010;12(14-15):1091-101.
12. Armstrong JA, Hart PD. Response of cultured macrophages to Mycobacterium tuberculosis, with observations on fusion of lysosomes with phagosomes. J Exp Med. 1971;134(3 Pt 1):713-40.
13. Russell DG. Mycobacterium tuberculosis: here today, and here tomorrow. Nat Rev Mol Cell Biol. 2001;2(8):569-77.
14. MacMicking JD, North RJ, LaCourse R, Mudgett JS, Shah SK, Nathan CF. Identification of nitric oxide synthase as a protective locus against tuberculosis. Proc Natl Acad Sci U S A. 1997;94(10):5243-8.
15. MacMicking JD, Nathan C, Hom G, Chartrain N, Fletcher DS, Trumbauer M, et al. Altered responses to bacterial infection and endotoxic shock in mice lacking inducible nitric oxide synthase. Cell. 1995;81(4):641-50.
16. Cooper AM, Segal BH, Frank AA, Holland SM, Orme IM. Transient loss of resistance to pulmonary tuberculosis in p47(phox-/-) mice. Infect Immun. 2000;68(3):1231-4.

17. Edwards KM, Cynamon MH, Voladri RK, Hager CC, DeStefano MS, Tham KT, et al. Iron-cofactored superoxide dismutase inhibits host responses to Mycobacterium tuberculosis. Am J Respir Crit Care Med. 2001;164(12):2213-9.
18. Via LE, Lin PL, Ray SM, Carrillo J, Allen SS, Eum SY, et al. Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. Infect Immun. 2008;76(6):2333-40.
19. Wayne LG, Hayes LG. An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence. Infect Immun. 1996;64(6):2062-9.
20. Wayne LG, Lin KY. Glyoxylate metabolism and adaptation of Mycobacterium tuberculosis to survival under anaerobic conditions. Infect Immun. 1982;37(3):1042-9.
21. Wayne LG. Dormancy of Mycobacterium tuberculosis and latency of disease. Eur J Clin Microbiol Infect Dis. 1994;13(11):908-14.
22. Sambandamurthy VK, Wang X, Chen B, Russell RG, Derrick S, Collins FM, et al. A pantothenate auxotroph of Mycobacterium tuberculosis is highly attenuated and protects mice against tuberculosis. Nat Med. 2002;8(10):1171-4.
23. Garcia-Garcia T, Poncet S, Derouiche A, Shi L, Mijakovic I, Noirot-Gros MF. Role of Protein Phosphorylation in the Regulation of Cell Cycle and DNA-Related Processes in Bacteria. Front Microbiol. 2016;7:184.
24. Parish T. Two-Component Regulatory Systems of Mycobacteria. Microbiol Spectr. 2014;2(1):MGM2-0010-2013.
25. Ashby MK. Survey of the number of two-component response regulator genes in the complete and annotated genome sequences of prokaryotes. FEMS Microbiol Lett. 2004;231(2):277-81.
26. Abramovitch RB, Rohde KH, Hsu FF, Russell DG. aprABC: a Mycobacterium tuberculosis complex-specific locus that modulates pH-driven adaptation to the macrophage phagosome. Mol Microbiol. 2011;80(3):678-94.
27. Pang X, Vu P, Byrd TF, Ghanny S, Soteropoulos P, Mukamolova GV, et al. Evidence for complex interactions of stress-associated regulons in an mprAB deletion mutant of Mycobacterium tuberculosis. Microbiology (Reading). 2007;153(Pt 4):1229-42.
28. Glover RT, Kriakov J, Garforth SJ, Baughn AD, Jacobs WR, Jr. The two-component regulatory system senX3-regX3 regulates phosphate-dependent gene expression in Mycobacterium smegmatis. J Bacteriol. 2007;189(15):5495-503.
29. Haydel SE, Malhotra V, Cornelison GL, Clark-Curtiss JE. The prrAB two-component system is essential for Mycobacterium tuberculosis viability and is induced under nitrogen-limiting conditions. J Bacteriol. 2012;194(2):354-61.
30. Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K. Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. Mol Microbiol. 2002;43(3):717-31.
31. Park HD, Guinn KM, Harrell MI, Liao R, Voskuil MI, Tompa M, et al. Rv3133c/dosR is a transcription factor that mediates the hypoxic response of Mycobacterium tuberculosis. Mol Microbiol. 2003;48(3):833-43.
32. Ludwiczak P, Gilleron M, Bordat Y, Martin C, Gicquel B, Puzo G. Mycobacterium tuberculosis phoP mutant: lipoarabinomannan molecular structure. Microbiology (Reading). 2002;148(Pt 10):3029-37.

33. Chesne-Seck ML, Barilone N, Boudou F, Gonzalo Asensio J, Kolattukudy PE, Martin C, et al. A point mutation in the two-component regulator PhoP-PhoR accounts for the absence of polyketide-derived acyltrehaloses but not that of phthiocerol dimycocerosates in Mycobacterium tuberculosis H37Ra. J Bacteriol. 2008;190(4):1329-34.
34. Frigui W, Bottai D, Majlessi L, Monot M, Josselin E, Brodin P, et al. Control of M. tuberculosis ESAT-6 secretion and specific T cell recognition by PhoP. PLoS Pathog. 2008;4(2):e33.
35. Lee JS, Krause R, Schreiber J, Mollenkopf HJ, Kowall J, Stein R, et al. Mutation in the transcriptional regulator PhoP contributes to avirulence of Mycobacterium tuberculosis H37Ra strain. Cell Host Microbe. 2008;3(2):97-103.
36. Zahrt TC, Deretic V. Mycobacterium tuberculosis signal transduction system required for persistent infections. Proc Natl Acad Sci U S A. 2001;98(22):12706-11.
37. Ewann F, Jackson M, Pethe K, Cooper A, Mielcarek N, Ensergueix D, et al. Transient requirement of the PrrA-PrrB two-component system for early intracellular multiplication of Mycobacterium tuberculosis. Infect Immun. 2002;70(5):2256-63.
38. Gassel M, Siebers A, Epstein W, Altendorf K. Assembly of the Kdp complex, the multi-subunit K+-transport ATPase of Escherichia coli. Biochim Biophys Acta. 1998;1415(1):77-84.
39. Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, et al. A global protein kinase and phosphatase interaction network in yeast. Science. 2010;328(5981):1043-6.
40. Cohen P. Protein kinases--the major drug targets of the twenty-first century? Nat Rev Drug Discov. 2002;1(4):309-15.
41. Bhullar KS, Lagaron NO, McGowan EM, Parmar I, Jha A, Hubbard BP, et al. Kinase-targeted cancer therapies: progress, challenges and future directions. Mol Cancer. 2018;17(1):48.
42. Macek B, Gnad F, Soufi B, Kumar C, Olsen JV, Mijakovic I, et al. Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. Mol Cell Proteomics. 2008;7(2):299-307.
43. Macek B, Mijakovic I, Olsen JV, Gnad F, Kumar C, Jensen PR, et al. The serine/threonine/tyrosine phosphoproteome of the model bacterium Bacillus subtilis. Mol Cell Proteomics. 2007;6(4):697-707.
44. Sun X, Ge F, Xiao CL, Yin XF, Ge R, Zhang LH, et al. Phosphoproteomic analysis reveals the multiple roles of phosphorylation in pathogenic bacterium Streptococcus pneumoniae. J Proteome Res. 2010;9(1):275-82.
45. Ravichandran A, Sugiyama N, Tomita M, Swarup S, Ishihama Y. Ser/Thr/Tyr phosphoproteome analysis of pathogenic and non-pathogenic Pseudomonas species. Proteomics. 2009;9(10):2764-75.
46. Prisic S, Dankwa S, Schwartz D, Chou MF, Locasale JW, Kang CM, et al. Extensive phosphorylation with overlapping specificity by Mycobacterium tuberculosis serine/threonine protein kinases. Proc Natl Acad Sci U S A. 2010;107(16):7521-6.
47. Potel CM, Lin MH, Heck AJR, Lemeer S. Defeating Major Contaminants in Fe(3+)-Immobilized Metal Ion Affinity Chromatography (IMAC) Phosphopeptide Enrichment. Mol Cell Proteomics. 2018;17(5):1028-34.
48. Av-Gay Y, Everett M. The eukaryotic-like Ser/Thr protein kinases of Mycobacterium tuberculosis. Trends Microbiol. 2000;8(5):238-44.

49. Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J. 1995;9(8):576-96.
50. Galperin MY, Higdon R, Kolker E. Interplay of heritage and habitat in the distribution of bacterial signal transduction systems. Mol Biosyst. 2010;6(4):721-8.
51. Schneiker S, Perlova O, Kaiser O, Gerth K, Alici A, Altmeyer MO, et al. Complete genome sequence of the myxobacterium Sorangium cellulosum. Nat Biotechnol. 2007;25(11):1281-9.
52. Kannan N, Taylor SS, Zhai Y, Venter JC, Manning G. Structural and functional diversity of the microbial kinome. PLoS Biol. 2007;5(3):e17.
53. Cluzel ME, Zanella-Cleon I, Cozzone AJ, Futterer K, Duclos B, Molle V. The Staphylococcus aureus autoinducer-2 synthase LuxS is regulated by Ser/Thr phosphorylation. J Bacteriol. 2010;192(23):6295-301.
54. Bugrysheva J, Froehlich BJ, Freiberg JA, Scott JR. Serine/threonine protein kinase Stk is required for virulence, stress response, and penicillin tolerance in Streptococcus pyogenes. Infect Immun. 2011;79(10):4201-9.
55. Fleurie A, Cluzel C, Guiral S, Freton C, Galisson F, Zanella-Cleon I, et al. Mutational dissection of the S/T-kinase StkP reveals crucial roles in cell division of Streptococcus pneumoniae. Mol Microbiol. 2012;83(4):746-58.
56. Hatzios SK, Baer CE, Rustad TR, Siegrist MS, Pang JM, Ortega C, et al. Osmosensory signaling in Mycobacterium tuberculosis mediated by a eukaryotic-like Ser/Thr protein kinase. Proc Natl Acad Sci U S A. 2013;110(52):E5069-77.
57. Ortega C, Liao R, Anderson LN, Rustad T, Ollodart AR, Wright AT, et al. Mycobacterium tuberculosis Ser/Thr protein kinase B mediates an oxygen-dependent replication switch. PLoS Biol. 2014;12(1):e1001746.
58. Zheng X, Papavinasasundaram KG, Av-Gay Y. Novel substrates of Mycobacterium tuberculosis PknH Ser/Thr kinase. Biochem Biophys Res Commun. 2007;355(1):162-8.
59. Rieck B, Degiacomi G, Zimmermann M, Cascioferro A, Boldrin F, Lazar-Adler NR, et al. PknG senses amino acid availability to control metabolism and virulence of Mycobacterium tuberculosis. PLoS Pathog. 2017;13(5):e1006399.
60. Gil M, Lima A, Rivera B, Rossello J, Urdaniz E, Cascioferro A, et al. New substrates and interactors of the mycobacterial Serine/Threonine protein kinase PknG identified by a tailored interactomic approach. J Proteomics. 2019;192:321-33.
61. Cowley S, Ko M, Pick N, Chow R, Downing KJ, Gordhan BG, et al. The Mycobacterium tuberculosis protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo. Mol Microbiol. 2004;52(6):1691-702.
62. Nakedi KC, Calder B, Banerjee M, Giddey A, Nel AJM, Garnett S, et al. Identification of Novel Physiological Substrates of Mycobacterium bovis BCG Protein Kinase G (PknG) by Label-free Quantitative Phosphoproteomics. Mol Cell Proteomics. 2018;17(7):1365-77.
63. Kumar P, Kumar D, Parikh A, Rananaware D, Gupta M, Singh Y, et al. The Mycobacterium tuberculosis protein kinase K modulates activation of transcription from the promoter of mycobacterial monooxygenase operon through phosphorylation of the transcriptional regulator VirS. J Biol Chem. 2009;284(17):11090-9.
64. Cohen-Gonsaud M, Barthe P, Canova MJ, Stagier-Simon C, Kremer L, Roumestand C, et al. The Mycobacterium tuberculosis Ser/Thr kinase substrate Rv2175c is a DNA-binding protein regulated by phosphorylation. J Biol Chem. 2009;284(29):19290-300.

65. Canova MJ, Veyron-Churlet R, Zanella-Cleon I, Cohen-Gonsaud M, Cozzone AJ, Becchi M, et al. The Mycobacterium tuberculosis serine/threonine kinase PknL phosphorylates Rv2175c: mass spectrometric profiling of the activation loop phosphorylation sites and their role in the recruitment of Rv2175c. Proteomics. 2008;8(3):521-33.
66. Turapov O, Forti F, Kadhim B, Ghisotti D, Sassine J, Straatman-Iwanowska A, et al. Two Faces of CwlM, an Essential PknB Substrate, in Mycobacterium tuberculosis. Cell Rep. 2018;25(1):57-67 e5.
67. Alqaseer K, Turapov O, Barthe P, Jagatia H, De Visch A, Roumestand C, et al. Protein kinase B controls Mycobacterium tuberculosis growth via phosphorylation of the transcriptional regulator Lsr2 at threonine 112. Mol Microbiol. 2019;112(6):1847-62.
68. van Kessel JC, Hatfull GF. Recombineering in Mycobacterium tuberculosis. Nat Methods. 2007;4(2):147-52.
69. Rock JM, Hopkins FF, Chavez A, Diallo M, Chase MR, Gerrick ER, et al. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. Nat Microbiol. 2017;2:16274.
70. Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T, et al. Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. Proteomics. 2011;11(10):2019-26.
71. Mertins P, Tang LC, Krug K, Clark DJ, Gritsenko MA, Chen L, et al. Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. Nat Protoc. 2018;13(7):1632-61.
72. Zeng J, Platig J, Cheng TY, Ahmed S, Skaf Y, Potluri LP, et al. Protein kinases PknA and PknB independently and coordinately regulate essential Mycobacterium tuberculosis physiologies and antimicrobial susceptibility. PLoS Pathog. 2020;16(4):e1008452.
73. Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, et al. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. Cell. 2016;166(3):755-65.
74. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods. 2007;4(3):207-14.
75. Qian WJ, Liu T, Monroe ME, Strittmatter EF, Jacobs JM, Kangas LJ, et al. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. J Proteome Res. 2005;4(1):53-62.
76. Kim S, Gupta N, Pevzner PA. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. J Proteome Res. 2008;7(8):3354-63.
77. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367-72.
78. Clair G, Piehowski PD, Nicola T, Kitzmiller JA, Huang EL, Zink EM, et al. Spatially-Resolved Proteomics: Rapid Quantitative Analysis of Laser Capture Microdissected Alveolar Tissue Samples. Sci Rep. 2016;6:39223.
79. Shi T, Sun X, Gao Y, Fillmore TL, Schepmoes AA, Zhao R, et al. Targeted quantification of low ng/mL level proteins in human serum without immunoaffinity depletion. J Proteome Res. 2013;12(7):3353-61.

80. Shi T, Fillmore TL, Sun X, Zhao R, Schepmoes AA, Hossain M, et al. Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. Proc Natl Acad Sci U S A. 2012;109(38):15395-400.
81. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010;26(7):966-8.
82. He J, Sun X, Shi T, Schepmoes AA, Fillmore TL, Petyuk VA, et al. Antibody-independent targeted quantification of TMPRSS2-ERG fusion protein products in prostate cancer. Mol Oncol. 2014;8(7):1169-80.
83. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol. 2008;8:91.
84. Minch KJ, Rustad TR, Peterson EJ, Winkler J, Reiss DJ, Ma S, et al. The DNA-binding network of Mycobacterium tuberculosis. Nat Commun. 2015;6:5829.
85. Wu FL, Liu Y, Jiang HW, Luan YZ, Zhang HN, He X, et al. The Ser/Thr Protein Kinase Protein-Protein Interaction Map of M. tuberculosis. Mol Cell Proteomics. 2017;16(8):1491-506.
86. Cargnello M, Roux PP. Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. Microbiol Mol Biol Rev. 2011;75(1):50-83.
87. Bodenmiller B, Wanka S, Kraft C, Urban J, Campbell D, Pedrioli PG, et al. Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. Sci Signal. 2010;3(153):rs4.
88. Baer CE, Iavarone AT, Alber T, Sassetti CM. Biochemical and spatial coincidence in the provisional Ser/Thr protein kinase interaction network of Mycobacterium tuberculosis. J Biol Chem. 2014;289(30):20422-33.
89. Alber T. Signaling mechanisms of the Mycobacterium tuberculosis receptor Ser/Thr protein kinases. Curr Opin Struct Biol. 2009;19(6):650-7.
90. Mir M, Asong J, Li X, Cardot J, Boons GJ, Husson RN. The extracytoplasmic domain of the Mycobacterium tuberculosis Ser/Thr kinase PknB binds specific muropeptides and is required for PknB localization. PLoS Pathog. 2011;7(7):e1002182.
91. Soares NC, Spat P, Krug K, Macek B. Global dynamics of the Escherichia coli proteome and phosphoproteome during growth in minimal medium. J Proteome Res. 2013;12(6):2611-21.
92. Arnvig KB, Comas I, Thomson NR, Houghton J, Boshoff HI, Croucher NJ, et al. Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis. PLoS Pathog. 2011;7(11):e1002342.
93. Kelkar DS, Kumar D, Kumar P, Balakrishnan L, Muthusamy B, Yadav AK, et al. Proteogenomic analysis of Mycobacterium tuberculosis by high resolution mass spectrometry. Mol Cell Proteomics. 2011;10(12):M111 011627.
94. Kusebauch U, Ortega C, Ollodart A, Rogers RS, Sherman DR, Moritz RL, et al. Mycobacterium tuberculosis supports protein tyrosine phosphorylation. Proc Natl Acad Sci U S A. 2014;111(25):9265-70.
95. Sharma K, D'Souza RC, Tyanova S, Schaab C, Wisniewski JR, Cox J, et al. Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. Cell Rep. 2014;8(5):1583-94.

96. Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sassetti CM. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog. 2011;7(9):e1002251.
97. Carette X, Platig J, Young DC, Helmel M, Young AT, Wang Z, et al. Multisystem Analysis of Mycobacterium tuberculosis Reveals Kinase-Dependent Remodeling of the Pathogen-Environment Interface. mBio. 2018;9(2).
98. Fortuin S, Tomazella GG, Nagaraj N, Sampson SL, Gey van Pittius NC, Soares NC, et al. Phosphoproteomics analysis of a clinical Mycobacterium tuberculosis Beijing isolate: expanding the mycobacterial phosphoproteome catalog. Front Microbiol. 2015;6:6.
99. Verma R, Pinto SM, Patil AH, Advani J, Subba P, Kumar M, et al. Quantitative Proteomic and Phosphoproteomic Analysis of H37Ra and H37Rv Strains of Mycobacterium tuberculosis. J Proteome Res. 2017;16(4):1632-45.
100. Amanchy R, Periaswamy B, Mathivanan S, Reddy R, Tattikota SG, Pandey A. A curated compendium of phosphorylation motifs. Nat Biotechnol. 2007;25(3):285-6.
101. Pearson RB, Kemp BE. Protein kinase phosphorylation site sequences and consensus specificity motifs: tabulations. Methods Enzymol. 1991;200:62-81.
102. Marin O, Bustos VH, Cesaro L, Meggio F, Pagano MA, Antonelli M, et al. A noncanonical sequence phosphorylated by casein kinase 1 in beta-catenin may play a role in casein kinase 1 targeting of important signaling proteins. Proc Natl Acad Sci U S A. 2003;100(18):10193-200.
103. Friedmann M, Nissen MS, Hoover DS, Reeves R, Magnuson NS. Characterization of the proto-oncogene pim-1: kinase activity and substrate recognition sequence. Arch Biochem Biophys. 1992;298(2):594-601.
104. Greenstein AE, Echols N, Lombana TN, King DS, Alber T. Allosteric activation by dimerization of the PknD receptor Ser/Thr protein kinase from Mycobacterium tuberculosis. J Biol Chem. 2007;282(15):11427-35.
105. Roumestand C, Leiba J, Galophe N, Margeat E, Padilla A, Bessin Y, et al. Structural insight into the Mycobacterium tuberculosis Rv0020c protein and its interaction with the PknB kinase. Structure. 2011;19(10):1525-34.
106. Thakur M, Chaba R, Mondal AK, Chakraborti PK. Interdomain interaction reconstitutes the functionality of PknA, a eukaryotic type Ser/Thr kinase from Mycobacterium tuberculosis. J Biol Chem. 2008;283(12):8023-33.
107. Mittal E, Skowyra ML, Uwase G, Tinaztepe E, Mehra A, Koster S, et al. Mycobacterium tuberculosis Type VII Secretion System Effectors Differentially Impact the ESCRT Endomembrane Damage Response. mBio. 2018;9(6).
108. Prust N, van der Laarse S, van den Toorn HWP, van Sorge NM, Lemeer S. In-Depth Characterization of the Staphylococcus aureus Phosphoproteome Reveals New Targets of Stk1. Mol Cell Proteomics. 2021;20:100034.
109. Sajid A, Arora G, Gupta M, Singhal A, Chakraborty K, Nandicoori VK, et al. Interaction of Mycobacterium tuberculosis elongation factor Tu with GTP is regulated by phosphorylation. J Bacteriol. 2011;193(19):5347-58.
110. Gupta M, Sajid A, Arora G, Tandon V, Singh Y. Forkhead-associated domain-containing protein Rv0019c and polyketide-associated protein PapA5, from substrates of serine/threonine protein kinase PknB to interacting proteins of Mycobacterium tuberculosis. J Biol Chem. 2009;284(50):34723-34.

111. Leiba J, Syson K, Baronian G, Zanella-Cleon I, Kalscheuer R, Kremer L, et al. Mycobacterium tuberculosis maltosyltransferase GlgE, a genetically validated antituberculosis target, is negatively regulated by Ser/Thr phosphorylation. J Biol Chem. 2013;288(23):16546-56.
112. Gee CL, Papavinasasundaram KG, Blair SR, Baer CE, Falick AM, King DS, et al. A phosphorylated pseudokinase complex controls cell wall synthesis in mycobacteria. Sci Signal. 2012;5(208):ra7.
113. Sajid A, Arora G, Gupta M, Upadhyay S, Nandicoori VK, Singh Y. Phosphorylation of Mycobacterium tuberculosis Ser/Thr phosphatase by PknA and PknB. PLoS One. 2011;6(3):e17871.
114. Park ST, Kang CM, Husson RN. Regulation of the SigH stress response regulon by an essential protein kinase in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 2008;105(35):13105-10.
115. Greenstein AE, MacGurn JA, Baer CE, Falick AM, Cox JS, Alber T. M. tuberculosis Ser/Thr protein kinase D phosphorylates an anti-anti-sigma factor homolog. PLoS Pathog. 2007;3(4):e49.
116. Molle V, Soulat D, Jault JM, Grangeasse C, Cozzone AJ, Prost JF. Two FHA domains on an ABC transporter, Rv1747, mediate its phosphorylation by PknF, a Ser/Thr protein kinase from Mycobacterium tuberculosis. FEMS Microbiol Lett. 2004;234(2):215-23.
117. Canova MJ, Kremer L, Molle V. The Mycobacterium tuberculosis GroEL1 chaperone is a substrate of Ser/Thr protein kinases. J Bacteriol. 2009;191(8):2876-83.
118. Veyron-Churlet R, Molle V, Taylor RC, Brown AK, Besra GS, Zanella-Cleon I, et al. The Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein synthase III activity is inhibited by phosphorylation on a single threonine residue. J Biol Chem. 2009;284(10):6414-24.
119. Molle V, Brown AK, Besra GS, Cozzone AJ, Kremer L. The condensing activities of the Mycobacterium tuberculosis type II fatty acid synthase are differentially regulated by phosphorylation. J Biol Chem. 2006;281(40):30094-103.
120. Corrales RM, Molle V, Leiba J, Mourey L, de Chastellier C, Kremer L. Phosphorylation of mycobacterial PcaA inhibits mycolic acid cyclopropanation: consequences for intracellular survival and for phagosome maturation block. J Biol Chem. 2012;287(31):26187-99.
121. Molle V, Kremer L, Girard-Blanc C, Besra GS, Cozzone AJ, Prost JF. An FHA phosphoprotein recognition domain mediates protein EmbR phosphorylation by PknH, a Ser/Thr protein kinase from Mycobacterium tuberculosis. Biochemistry. 2003;42(51):15300-9.
122. Jang J, Stella A, Boudou F, Levillain F, Darthuy E, Vaubourgeix J, et al. Functional characterization of the Mycobacterium tuberculosis serine/threonine kinase PknJ. Microbiology (Reading). 2010;156(Pt 6):1619-31.
123. Arora G, Sajid A, Gupta M, Bhaduri A, Kumar P, Basu-Modak S, et al. Understanding the role of PknJ in Mycobacterium tuberculosis: biochemical characterization and identification of novel substrate pyruvate kinase A. PLoS One. 2010;5(5):e10772.
124. Piehowski PD, Petyuk VA, Orton DJ, Xie F, Moore RJ, Ramirez-Restrepo M, et al. Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis. J Proteome Res. 2013;12(5):2128-37.
125. Anderson D, Koch CA, Grey L, Ellis C, Moran MF, Pawson T. Binding of SH2 domains of phospholipase C gamma 1, GAP, and Src to activated growth factor receptors. Science. 1990;250(4983):979-82.

126. Ottinger EA, Botfield MC, Shoelson SE. Tandem SH2 domains confer high specificity in tyrosine kinase signaling. J Biol Chem. 1998;273(2):729-35.
127. Ben-Levy R, Leighton IA, Doza YN, Attwood P, Morrice N, Marshall CJ, et al. Identification of novel phosphorylation sites required for activation of MAPKAP kinase-2. EMBO J. 1995;14(23):5920-30.
128. Markevich NI, Hoek JB, Kholodenko BN. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. J Cell Biol. 2004;164(3):353-9.
129. Thomson M, Gunawardena J. Unlimited multistability in multisite phosphorylation systems. Nature. 2009;460(7252):274-7.
130. Weller M. Protein phosphorylation: Pion Ltd.; 1979.
131. Landry CR, Levy ED, Michnick SW. Weak functional constraints on phosphoproteomes. Trends Genet. 2009;25(5):193-7.
132. Ochoa D, Jarnuczak AF, Vieitez C, Gehre M, Soucheray M, Mateus A, et al. The functional landscape of the human phosphoproteome. Nat Biotechnol. 2020;38(3):365-73.
133. Browning DF, Busby SJ. Local and global regulation of transcription initiation in bacteria. Nat Rev Microbiol. 2016;14(10):638-50.
134. Martinez-Antonio A, Collado-Vides J. Identifying global regulators in transcriptional regulatory networks in bacteria. Curr Opin Microbiol. 2003;6(5):482-9.
135. Filtz TM, Vogel WK, Leid M. Regulation of transcription factor activity by interconnected post-translational modifications. Trends Pharmacol Sci. 2014;35(2):76-85.
136. Sakamoto Y, Yoshida M, Semba K, Hunter T. Inhibition of the DNA-binding and transcriptional repression activity of the Wilms' tumor gene product, WT1, by cAMP-dependent protein kinase-mediated phosphorylation of Ser-365 and Ser-393 in the zinc finger domain. Oncogene. 1997;15(17):2001-12.
137. Crabtree GR. Generic signals and specific outcomes: signaling through Ca2+, calcineurin, and NF-AT. Cell. 1999;96(5):611-4.
138. Shiama N. The p300/CBP family: integrating signals with transcription factors and chromatin. Trends Cell Biol. 1997;7(6):230-6.
139. Spencer E, Jiang J, Chen ZJ. Signal-induced ubiquitination of IkappaBalpha by the F-box protein Slimb/beta-TrCP. Genes Dev. 1999;13(3):284-94.
140. Lienhard GE. Non-functional phosphorylations? Trends Biochem Sci. 2008;33(8):351-2.
141. Rustad TR, Minch KJ, Ma S, Winkler JK, Hobbs S, Hickey M, et al. Mapping and manipulating the Mycobacterium tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network. Genome Biol. 2014;15(11):502.
142. Boshoff HI, Myers TG, Copp BR, McNeil MR, Wilson MA, Barry CE, 3rd. The transcriptional responses of Mycobacterium tuberculosis to inhibitors of metabolism: novel insights into drug mechanisms of action. J Biol Chem. 2004;279(38):40174-84.
143. Honaker RW, Leistikow RL, Bartek IL, Voskuil MI. Unique roles of DosT and DosS in DosR regulon induction and Mycobacterium tuberculosis dormancy. Infect Immun. 2009;77(8):3258-63.
144. O'Donnell G, Poeschl R, Zimhony O, Gunaratnam M, Moreira JB, Neidle S, et al. Bioactive pyridine-N-oxide disulfides from Allium stipitatum. J Nat Prod. 2009;72(3):360-5.
145. Raghavan S, Manzanillo P, Chan K, Dovey C, Cox JS. Secreted transcription factor controls Mycobacterium tuberculosis virulence. Nature. 2008;454(7205):717-21.

146. Rustad TR, Harrell MI, Liao R, Sherman DR. The enduring hypoxic response of Mycobacterium tuberculosis. PLoS One. 2008;3(1):e1502.
147. Chao JD, Papavinasasundaram KG, Zheng X, Chavez-Steenbock A, Wang X, Lee GQ, et al. Convergence of Ser/Thr and two-component signaling to coordinate expression of the dormancy regulon in Mycobacterium tuberculosis. J Biol Chem. 2010;285(38):29239-46.
148. Park EJ, Kwon YM, Lee JW, Kang HY, Oh JI. Dual control of RegX3 transcriptional activity by SenX3 and PknB. J Biol Chem. 2019;294(28):11023-34.
149. Perez J, Castaneda-Garcia A, Jenke-Kodama H, Muller R, Munoz-Dorado J. Eukaryotic-like protein kinases in the prokaryotes and the myxobacterial kinome. Proc Natl Acad Sci U S A. 2008;105(41):15950-5.