Identification of diverse targets of arginine phosphorylation in *Mycolicibacterium smegmatis* by shotgun proteomics

Emmanuel Ogbonna[1] and Karl R. Schmitz[1,2]*

Department of Biological Sciences[1]
Department of Chemistry & Biochemistry[2]
University of Delaware
Newark DE, 19716

*corresponding author: schmitzk@udel.edu

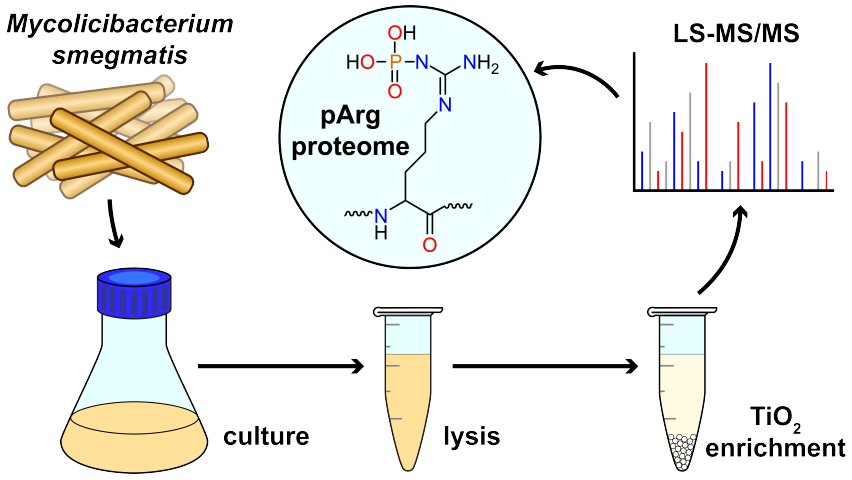
running title: Identification of arginine phosphorylation in *M. smegmatis*

abbreviations:

Abstract

Tuberculosis is a leading cause of worldwide infectious mortality**.** The prevalence of multidrug-resistant *Mycobacterium tuberculosis* (*Mtb*) infections drives an urgent need to exploit new drug targets. One such target is the ATP-dependent protease ClpC1P1P2, which is strictly essential for viability. However, few proteolytic substrates of mycobacterial ClpC1P1P2 have been identified to date. Recent studies in *Bacillus subtilis* have shown that the orthologous ClpCP protease recognizes proteolytic substrates bearing post-translational arginine phosphorylation. Several lines of evidence suggest that ClpC1P1P2 similarly recognizes phosphoarginine-bearing proteins, but the existence of phosphoarginine modifications in mycobacteria has remained in question. Here, we confirm the presence of post-translational phosphoarginine modifications in *Mycolicibacterium smegmatis* (*Msm*), a nonpathogenic surrogate of *Mtb.* Using a phosphopeptide enrichment workflow coupled with shotgun phosphoproteomics, we identify arginine phosphosites on a diverse collection of targets within the *Msm* proteome. Physicochemical and functional characterization of targets suggest that arginine phosphorylation is applied in a sequence-independent manner as part of a proteome-wide quality control pathway. Our findings provide new evidence supporting the existence of phosphoarginine-mediated proteolysis by ClpC1P1P2 in mycobacteria and other actinobacterial species.

GRAPHICAL ABSTRACT



INTRODUCTION

Tuberculosis is a leading cause of worldwide infectious mortality, ranking above HIV/AIDS and Ebola, and is one of the top ten leading causes of death overall (1). Advances in diagnosis, vaccinations and therapeutics have reduced tuberculosis morbidity and mortality. However, the prevalence of multidrug-resistance in the causative bacterium, *Mycobacterium tuberculosis* (*Mtb*), remains high, with over 200,000 multidrug-resistant tuberculosis infections identified in 2019 alone (1). These statistics underscore the urgent need to discover new drugs and exploit new molecular targets. One promising target is the mycobacterial Clp protease. Several studies have revealed that Clp protease components are strictly essential for *Mtb* viability (2-6), and have demonstrated that these enzymes are viable targets for anti-*Mtb* therapeutics (7-12).

Clp proteases are large complexes that mechanically unfold and destroy native cytosolic proteins (13, 14). They consist of a core barrel-shaped peptidase and an associated hexameric ATP-dependent unfoldase. The mycobacterial Clp peptidase is a heteromer, composed of distinct ClpP1 and ClpP2 isoforms that assemble into separate heptameric rings (15, 16). Although both isoforms possess serine peptidase active sites, each is individually inactive (16, 17). ClpP1 and ClpP2 rings stack face-to-face to create the catalytically active hetero-oligomeric tetradecamer, ClpP1P2 (15, 18, 19). ClpP1P2 assembly is stabilized by interactions from unfoldase and, *in vitro*, by some N-blocked peptides that bind to the peptidase active sites (16, 19). ClpC1 is one of two alternative mycobacterial unfoldases that can assemble with ClpP1P2 to form a functional protease. ClpC1 is an 848-residue protein with a globular N-terminal domain (NTD) and two AAA+ (ATPases associated with various cellular activities) modules. Ring-shaped hexamers of ClpC1 dock coaxially on the surface of the ClpP2 ring (20). ClpC1 serves a critical regulatory role in proteolysis, by selectively recognizing proteolytic substrates.

Clp proteases have become major targets of novel drug development against *Mtb* and other pathogenic bacteria. Several classes of antimicrobials specifically target the Clp peptidase,

including dysregulators (e.g., acyldepsipeptides (21, 22)), activators (e.g., sclerotiamide (23)), and catalytic inhibitors (e.g., β-lactones and boronate compounds (11, 24)). Importantly, multiple cyclic peptides have been discovered that exhibit anti-*Mtb* activity through dysregulation of ClpC1, including cyclomarin A, lassomycin, ecumicin and rufomycin (9, 10, 25, 26). These compounds bind to the ClpC1 NTD and disrupt proteolysis by uncoupling unfoldase activity from proteolysis, causing uncontrolled degradation of cellular proteins, and/or blocking substrate recognition. The antimicrobial activity ClpC1-targeting compounds underscores the importance of ClpC1P1P2 in mycobacteria. However, the specific proteolytic functions responsible for its essentially remain obscure, and only a handful of protein substrates have been identified (6, 27-33).

An expanded understanding of the physiological roles played by ClpC1P1P2 would bolster efforts to develop antimicrobial compounds. In other bacteria, Clp proteases participate in cellular processes ranging from targeted pathway regulation to cell-wide quality control programs. Multiple mechanisms of substrate recognition have been described, including direct interactions with the unfoldase (6, 31-33) and indirect recognition through proteolytic adaptors (27, 34). Recently, studies in *Bacillus subtilis* and related Firmicutes have demonstrated that post-translational arginine phosphorylation marks some proteins for destruction by ClpCP (35-37). Dual phosphoarginine (pArg) binding sites on the *B. subtilis* ClpC NTD allow ClpCP to recognize phosphoarginylated proteins as proteolytic substrates (36). Phosphoarginine-mediated proteolysis in these bacteria is upregulated during stress, through activation of the arginine kinase McsB (38). Phosphoproteomic studies have uncovered widespread arginine phosphorylation during heat stress, implicating ClpCP in turnover of misfolded proteins (36, 37). Additionally, McsB regulates the global stress response through targeted phosphorylation of the negative transcriptional regulators CtsR and HrcA (37-39). Proteolysis of these targets by ClpCP allows transcriptional activation of stress response genes (38). Interestingly, ClpC and ClpP are themselves targets of McsB, and specific pArg sites on ClpC are required for its activation by

McsB (37, 40). These studies underline the role of arginine phosphorylation in Firmicutes as both a degradation signal and a regulatory mechanism.

Several lines of evidence suggest that an analogous pArg-mediated proteolytic pathways exists in mycobacteria. Sequence and structural data reveal overall homology between the NTDs of *Mtb* ClpC1 and *B. subtilis* ClpC, as well as strong conservation of the residues surrounding the pArg-binding sites (36, 41). Moreover, *in vitro* experiments confirm that the *Mtb* ClpC1 NTD does indeed interact with both free pArg and with arginine-phosphorylated model substrates (41). However, the existence pArg-mediated proteolysis in mycobacteria remains in question, as no mycobacterial McsB homologs are known and phosphoarginine modifications have not yet been described in these bacteria.

Here, we confirm the existence of post-translational phosphoarginine modifications in *Mycolicibacterium smegmatis* (*Msm*), a nonpathogenic surrogate of *Mtb.* Using a phosphopeptide enrichment workflow coupled to shotgun phosphoproteomics, we identify arginine phosphosites on a diverse collection of targets within the *Msm* proteome. Physiochemical and functional characterization of targets suggest that arginine phosphorylation is part of a proteome-wide quality control pathway. Our findings provide new evidence supporting the existence of pArg-mediated proteolysis by ClpC1P1P2 in mycobacteria and other actinobacterial species.

EXPERIMENTAL PROCEDURES

ClpC1 Sequence Analysis — Actinobacterial homologs of *Msm* ClpC1 were identified using HMMER (42) and aligned using Clustal Omega (43). Fragmentary sequences were removed. To reduce overrepresentation of similar taxa (e.g., similar *Mtb* strains), alignments were pruned such that no two sequences had greater than 90% sequence identity. Positional conservation scores were exported from Jalview (44) and plotted as a heatmap in Prism (Graphpad). Sequence logo of the NTD was generated using WebLogo (45).

Protein purification and binding assays — A nucleotide sequence encoding the NTD of ClpC1 (codons 1 - 147) was amplified from *M. smegmatis* MC^2155 gDNA (ATCC) and cloned into a pET22b-derived vector (EMD Millipore) in frame with a C-terminal LPETGG sortase recognition sequence (46) and 6xHis tag. $ClpC1^{NTD}$ was overexpressed in *E. coli* strain ER2566 (NEB) by induction with 0.5 mM isopropyl β-D-1-thiogalactopyranoside at 30°C for 4 h. Cells were harvested by centrifugation, resuspended in lysis buffer (25 mM HEPES, 500 mM NaCl, 10 mM imidazole, 10% glycerol, pH 7.5), and lysed by sonication. $ClpC1^{NTD}$ was purified from clarified lysate by Ni-NTA (Marvelgent Biosciences) and anion exchange (Source Q, Cytiva) chromatography. $ClpC1^{NTD}$ was fluorescently labeled via sortase transpeptidation (46), by 2 h incubation with sortase A (~1/30 molar ratio) and a 2-fold molar excess of Gly-Gly-Asn-Lys-(fluorescein-isothiocyanate) peptide (Biomatik) in PBS buffer (pH 7.4). Excess peptide was removed by size exclusion chromatography (Superdex 75, Cytiva). Purified $ClpC1^{NTD\text{-}FITC}$ was concentrated and stored in CPD Buffer (25 mM HEPES, 200 mM NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, pH 7.0).

Binding of phosphoarginine (MilliporeSigma) to 0.1 μM purified $ClpC1^{NTD\text{-}FITC}$ was assayed in CPD (containing 0.05% Tween 20 and 8 mg/ml BSA) by microscale thermophoresis using a Monolith NT.115 (NanoTemper). Thermophoretic data were fit to a quadratic single site binding equation in Prism (Graphpad).

Cell Culture Conditions — Liquid starter cultures of *M. smegmatis* MC^2155 were prepared in 10 mL Middlebrook broth base (HiMedia) containing 0.2% w/v glucose (TCI), 0.2% v/v glycerol (Fisher Scientific), 0.05% v/v Tween 80, and grown for 60 hours at 37 ºC with orbital shaking. Saturated starter cultures were sub-cultured into 200 mL of fresh media (at a starting A_{600} of 0.05), and grown at 37 ºC until late log phase or early stationary phase (A_{600} 0.6-1.0). Cells were harvested by centrifugation at 9,000 × g for 20 min at 4 ºC, and pellets were resuspended in 5 mL of ice-cold lysis buffer (25 mM HEPES, 200 mM KCl, 10 mM $MgCl_2$, 0.1 mM EDTA, pH 7.5) containing 10 mM ATP (Fisher), 200 μL phosphatase inhibitor cocktail (MilliporeSigma), 3 mM

NaF (Fisher), 20 mM Na_3VO_4 (Acros Organics), 50 µL protease inhibitor cocktail (Thermo Fisher), and 0.2 mM phenylmethylsulfonyl fluoride (Acros Organics), and 200 mM dithiotreitol (DTT; Fisher). Cells were lysed at high pressure using a microfluidizer (Microfluidics). Lysates were clarified at 16,000 × g for 30 min at 4 °C, and supernatant was stored at -80 °C prior to further processing. Total protein content was estimated by Bradford assay (Bio-Rad).

Filter-Aided Sample Preparation for Mass Spectrometry — 3 mL of clarified cell lysates were prepared for mass spectrometry using filter-aided sample preparation (FASP) (37, 47). Samples were incubated at 56 °C for 50 min for complete reduction of disulfide bonds by DTT. Afterwards, samples were diluted in 12 mL 0.1 M Tris (Gold Biotechnology), pH 7.5, containing 8 M urea and 55 mM 2-iodoacetamide (both from Acros Organics) in an Amicon filter (30,000 MWCO, MilliporeSigma) and incubated in the dark at room temperature for 45 min to carboxymethylate free cysteines. After alkylation, samples were washed twice with 15 mL 0.1 M Tris, pH 7.5, containing 8 M urea (by centrifugation at 4,000 × g), then washed twice with 15 mL 0.1 M Tris, pH 7.5, containing 4 M urea, and finally twice with 15 mL 100 mM ammonium bicarbonate (Honeywell). After the last wash step, retentate was reduced to 2 mL by spin concentration.

In-Solution Trypsin Proteolytic Cleavage — MS Grade Trypsin Protease (Pierce) was dissolved to obtain a 1 µg/µL stock in 100 mM ammonium bicarbonate. Trypsin digestion of processed *M. smegmatis* lysates was performed in an Amicon filter unit at a 100:1 protein:trypsin ratio overnight at 37 °C. Upon digestion, peptides were recovered by centrifugation at 4,000 × g for 10 minutes, followed by the addition of 50 µL of 0.5 M NaCl, and a second round of centrifugation. Eluted samples were dried using a vacuum centrifuge. To remove residual ammonium bicarbonate and NaCl, samples were desalted using a HyperSep C18 column (Thermo Scientific).

TiO_2 Enrichment of Phosphopeptides — Phosphopeptides were enriched from tryptic peptide samples were using Titansphere TiO_2 beads (5 µm; GL Sciences) in mildly acidic buffer conditions, to limit acid hydrolysis of phosphoarginine (36, 37). 5 mg TiO_2 resin was resuspended in 1 mL Binding Buffer (300 mg/mL lactic acid, 12.5% acetic acid, 60% acetonitrile, 0.2% heptafluorobutyric acid, pH 4 with NH_3). Lyophilized peptide samples were re-dissolved in the TiO_2 suspension, incubated for 45 min at 20 ^oC with gentle agitation, then transferred to graphite spin columns (Thermo Fisher). Unbound peptides were removed by a first wash step with 500 µL Binding Buffer at 1500 × g, followed by three rapid wash steps using Wash Solution A (200 mg/ml lactic acid, 75% acetonitrile, 2% trifluoroacetic acid, 2% heptafluorobutyric acid), Wash Solution B (200 mg/mL lactic acid, 75% acetonitrile, 10% acetic acid, 0.1% heptafluorobutyric acid, pH 4 with NH_3), Wash Solution C (80% acetonitrile, 10% acetic acid), respectively. The resin was then incubated twice with 100 µL Elution Solution (1% NH_3, 30 mM ammonium phosphate) for 15 minutes, and eluate containing phosphopeptides was collected by centrifugation. Samples were lyophilized and stored at -80 ^oC prior to mass spectrometry.

Liquid chromatography - tandem mass spectrometry (LC-MS/MS) — Lyophilized samples were resuspended in 0.5% acetic acid, pH 4.5. Samples were then analyzed on a Q Exactive Orbitrap interfaced with Ultimate 3000 Nano-LC system (Thermo Fisher Scientific). Analysis by Q Exactive Orbitrap was as follows: 10 µL of trypsin-digested sample was loaded on an Acclaim PepMap RSLC column (75 µm x 15 cm nanoViper) via an autosampler. Solvent A consisted of water with 0.1% formic acid and Solvent B was acetonitrile containing 0.1 % formic acid. Samples were analyzed using a 150-min gradient running from 2 % Solvent B to 95 % Solvent B at a flow rate of 0.3 µL/min. MS data was acquired using a data-dependent top10 method that dynamically selected the most abundant precursor ions from the survey scan for high energy collision-induced dissociation (HCD) fragmentation using a stepped normalized collision energy of 28, 30, and 35

eV. Survey scans were acquired at a resolution of 70,000 at m/z 200 on the Q Exactive. LS-MS/MS data was collected in independent biological triplicates.

Mass Spectrometry Data Analysis —Proteomic analysis was performed in the Proteome Discoverer software suite (version 1.4, Thermo Fisher). Raw data was searched against the *M. smegmatis* (strain ATCC 700084 / MC^2 155) UniProt Reference Proteome (Proteome ID UP000000757; 6,602 entries in total) using Sequest HT (University of Washington and Thermo Fisher) (48). Iodoacetamide-mediated cysteine carboxymethylation was set as a static modification, phosphorylation of arginine residues was entered as a dynamic modification, and partial trypsinization - with maximum of two missed cleavages - was allowed. Precursor mass tolerance was set at 10 ppm, while allowing fragment ions to have a mass deviation of 0.8 Da for the HCD data. Phosphopeptide hits were re-analyzed using the PhosphoRS node, and only modifications with a score ≥75% were selected. C-terminal phosphoarginine hits were excluded, and only proteins with positive Sequest HT total ion scores were considered for further analysis. Following these selection criteria, 328 unique phosphoarginine modifications were found on 222 protein, out of 667 proteins identified by MS/MS (**Table 1; Supplemental Table S1**).

Gene Ontology Annotation Analysis — Gene Ontology (GO) annotation was performed using the OmicsBox software suite (49). Homologs were identified by BLAST against the SwissProt/UniProt database (50), with a minimum expectation value 10^{-3}. Homolog annotations were compiled from the InterPro database (51). BLAST and InterPro results were used to generate GO terms in terms of biological process.

Sequence and Physicochemical Analysis of Arginine Phosphosites — For structural analysis, proteins were processed by the MFDp2 server (52) to evaluate secondary structure (PSIPRED (53), positional disorder (MFDp2), and fractional sequence disorder (Discon (54)). Solvent

accessible surface area of arginine residues in published structures was calculated by PISA (**Supplemental Table S2**) (55). For physiochemical analysis, sequences were normalized to a consistent length by incorporated ten residues before and after each arginine, where possible (positions <10 residues from a terminus resulted in shorter sequences). For comparison, control datasets were constructed consisting of i) all arginines present in the 667 proteins identified by MS/MS (21,557 total) or ii) all arginine present in the *M. smegmatis* proteome (151,303 total). The average hydrophobicity and the average charge at pH 7.0 of residues flaking the central arginine was calculated using the Kyte-Doolittle scale (56) and published pKa values (57), respectively. Sequence logos of aligned 21-residue segments centered on phosphoarginines and control arginine datasets were generated using WebLogo (45).

Data Availability — The mass spectrometry data from this publication have been submitted to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org), and assigned the identifier _____.

RESULTS

Phosphoarginine Binding Sites in the ClpC1 NTD Are Conserved Across Actinobacteria – ClpC1 is an essential enzyme in *Mtb* and *Msm* (3, 4), yet few specific cellular roles or proteolytic substrates of the ClpC1P1P2 protease are known. To gain insight into its cellular function, we began by examining ClpC1 sequence conservation across the phylum Actinobacteria by constructing an alignment of 1,195 actinobacterial orthologs (**Fig. 1A**). Like most other type-II Clp unfoldases, ClpC1 possesses a ~150 amino acid N-terminal domain (NTD), which likely participates in regulation and substrate recognition, but lacks mechanical function or a direct role in catalysis (34, 58). Surprisingly, the NTD is as well or better conserved than the D1 and D2 AAA+ ATPase rings, suggesting that the ClpC1 NTD has important and conserved functions across Actinobacteria. This may help explain why naturally occurring antibiotics that target and dysregulate ClpC1 have evolved to specifically bind the NTD (9,10, 21, 25, 59-61).

In *Bacillus subtilis* and other Firmicutes, ClpCP recognizes phosphorylated arginine residues as degradation signals via twin binding sites on opposite ends of the ClpC NTD (36). Examination of actinobacterial ClpC1 sequences revealed that the residues known to be important for phosphoarginine binding in *B. subtilis* are conserved across actinobacterial orthologs (**Fig. 1B**). Additionally, we were able to model bound phosphoarginines in an existing structure of the *M. tuberculosis* ClpC1 NTD (62), with minor side chains rearrangements, based on the binding mode observed to the *B. subtilis* ClpC NTD (36) (**Fig. 1C,D**). Finally, we directly assessed binding of phosphoarginine to purified *Msm* ClpC1 NTD by microscale thermophoresis (**Fig. 1E**). Phosphoarginine bound with an affinity of 1.4 μM, similar to the ~5 μM affinity previously reported to the *Mtb* ClpC1 NTD (37). The ability of the ClpC1 NTD to bind pArg and the conservation of pArg binding modules across Actinobacteria provides strong indirect evidence for the existence of this post-translational modification in this phylum.

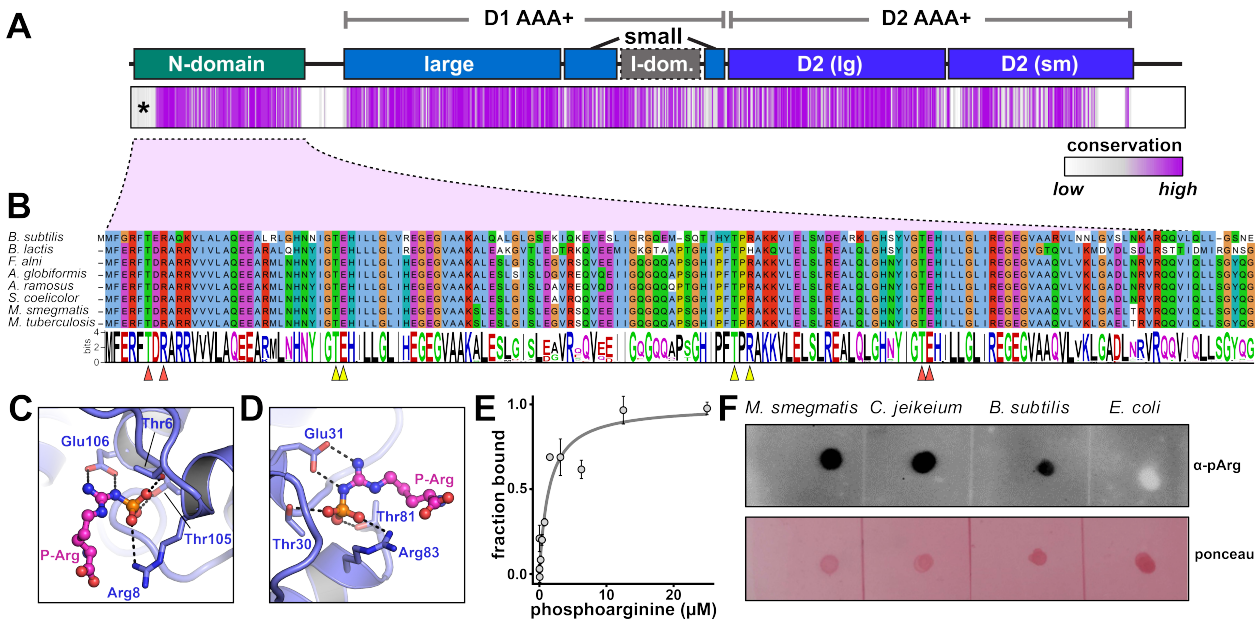


Figure 1. **Indirect evidence for phosphoarginine in actinobacteria**. **A**) The domain organization of ClpC1 is shown above a plot of amino acid conservation among actinobacterial ClpC1 orthologs, where conservation at each alignment position is plotted as a colored vertical strip. The apparent poor conservation at the beginning of the N-domain (marked by *) likely reflects mis-annotation of the start site in some entries. **B**) The sequence of the *Bacillus subtilis* ClpC N-domain is shown above equivalent regions from several actinobacterial ClpC orthologs. The sequence logo below shows amino acid conservation in actinobacterial ClpC1 orthologs. Arrows mark positions reported to be important for phosphoarginine binding to site 1 (orange) or site 2 (yellow) (41). Phosphoarginines from the crystal structure of *B. subtilis* ClpC N-domain (5HBN) were modeled on the putative *M. tuberculosis* ClpC1 N-domain (6PBQ) phosphoarginine-binding sites 1 (**C**) and 2 (**D**) (62). **E**) Binding of phosphoarginine to *Msm* ClpC1 NTD was measured by microscale thermophoresis. Data were fit to a noncooperative binding model (gray curve), yielding a K_D of 1.4 ± 0.6 μM. Values are averages of three technical replicates (N = 3) ± one SD. **F**) Dot blots of *Mycolicibacterium smegmatis*, *Corynebacterium jeikeium*, *Bacillus subtilis*, and *Escherichia coli* cell lysates, probed with anti-phosphoarginine antibody (**63**) or stained with Ponceau S.

Identification of Phosphoarginine Modifications in Mycolicibacterium smegmatis – To test for the physiological existence of arginine phosphorylation, we probed dot blots of several bacterial lysates with a phosphoarginine-specific antibody (**Fig. 1F**) (63). In line with prior studies, signal was observed in *B. subtilis* lysate (34), but not in lysate from *E. coli*, which lacks arginine phosphorylation (64, 65). We also observed a positive immunoblot reaction in lysates from *Msm* and a second actinobacterium, *Corynebacterium jeikeium*. These observations suggest that pArg modifications occur at least within the suborder Corynebacterineae, which encompasses *Corynebacterium*, *Mycolicibacterium*, and *Mycobacterium*.

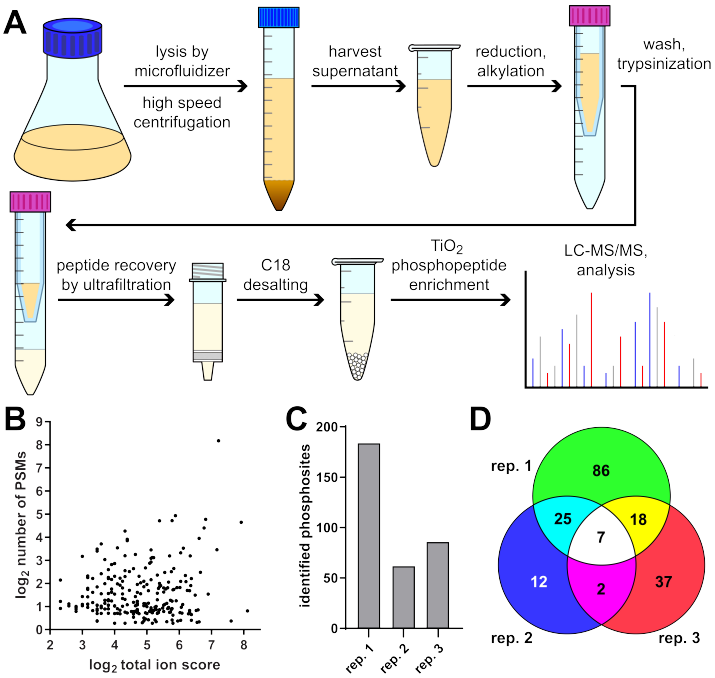


Figure 2: Identification of phosphoarginine modifications by mass spectrometry. A) Workflow for arginine phosphoproteomics in *Mycolicibacterium smegmatis*. **B**) Arginine-phosphorylated proteins plotted by number of peptide spectrum matches (PSMs) vs total ion score. **C**) Number of unique arginine phosphosites observed in each of three biological replicates. **D**) Venn diagram showing the number of arginine-phosphorylated proteins observed in multiple replicates.

To determine which specific cellular proteins carry pArg modifications in *Msm*, we employed an unbiased shotgun proteomics approach. Phosphoarginine is acid-labile and has a short half-life in the acidic TFA-containing solvent systems typically used for proteomic sample preparation and LS-MS/MS (37). Optimized protocols for arginine phosphoproteomics have been reported that utilize alternative solvent systems at pH ≥ 4 for most steps (36, 37). We adapted these methods to our workflow for sample preparation, phosphopeptide enrichment, and LC-MS/MS, thereby minimizing phosphoarginine hydrolysis (**Fig. 2A**). Prior phosphoarginine discovery efforts in *B. subtilis* and *S. aureus* utilized deletion strains lacking known arginine phosphatases, thus increasing the abundance of phosphoarginine in the samples (35, 37). However, no arginine phosphatase has yet been described in *Msm* or *Mtb*, and sequence searches reveal no conserved actinobacterial homologs of Firmicute arginine phosphatases, thus

our approach utilized wild-type *Msm*.

Msm cell lysates were subjected to a modified filter-aided sample preparation protocol and trypsinized in solution (47). Phosphopeptides were selectively enriched using TiO_2 resin (66). Three biological replicates were analyzed by LC-MS/MS using higher-energy collisional dissociation (HCD) fragmentation. Data analysis was performed using Sequest HT (48), and phosphosite localization was evaluated using the PhosphoRS node of the Proteome Discoverer software suite (67). To reduce the likelihood of incorrect assignment, additional threshold criteria were employed to validate arginine phosphosites: i) peptides with a C-terminal pArg were excluded because manual validation of the corresponding spectra was ambiguous, and because this modification is thought to physiologically impair trypsin digestion (37), ii) only proteins with positive Sequest HT total ion scores were considered for further analysis, and iii) only peptides with a phosphosite localization probability (PhosphoRS scores) ≥75% were selected (68). Using this workflow, we identified 328 unique phosphoarginine sites in 222 of the 667 cellular proteins identified overall (**Fig. 2B; Table 1**; **Supplemental Table S1**). Surprisingly, the list of specific phosphosites varied substantially across replicates, with only seven arginine-phosphorylated proteins observed across all three replicates (**Fig. 2C, D**). This variability suggests that arginine phosphorylation occurs sporadically on a large number of proteins, rather than on a discrete set of cellular targets.

Table 2: Proteins with 3 or more phosphoarginine modifications

Gene Name	Protein ID	Protein Name	Phosphosite	Functional Category
AceE	A0R0B0	Pyruvate dehydrogenase E1 component (PDH E1 component)	R27, R174, R313, R455, R811, R813	Redox Processes
MSMEG_4510	A0R0U4	Peptide synthetase mbtf	R200, R520, R970, R1009, R1069, R1098	Noncanonical Pathways
D806_012510	A0A2U9PKG0	Type III restriction enzyme, res subunit	R82, R256, R575, R738, R830, R831	General Nucleic acid-binding and Metabolism
MSMEG_4757	A0R1H7	Fatty acid synthase	R273, R362, R1197, R1342, R2610	Lipid Biosynthesis
XylB	A0QXD0	Xylulose kinase (Xylulokinase) (EC 2.7.1.17)	R127, R153, R206, R211	Carbohydrate metabolism
AftD	A0QPD4	Alpha-(1->3)-arabinofuranosyltransferase	R299, R301, R364	Carbohydrate metabolism

AlaS	A0QWQ4	Alanine--tRNA ligase (EC 6.1.1.7) (Alanyl-tRNA synthetase)	R710, R772, R864	General Nucleic acid-binding and Metabolism
HemB	A0QR20	Delta-aminolevulinic acid dehydratase (EC 4.2.1.24)	R19, R62, R63	Noncanonical Pathways
IleS	A0QX46	Isoleucine--tRNA ligase (EC 6.1.1.5) (Isoleucyl-tRNA synthetase)	R430, R652, R769	General Nucleic acid-binding and Metabolism
MSMEG_0848	A0QQR3	NADH ubiquinone oxidoreductase subunit 5	R4, R840, R851	Energy Metabolism
MSMEG_1941	A0QTR9	Helicase, UvrD/Rep family protein	R76, R391, R578	Stress
MSMEG_1963	A0QTU1	Putative transcriptional regulatory protein	R40, R165, R850	Transcriptional Regulation
MSMEG_4168	A0QZW0	Propionate CoA-transferase	R8, R49, R212	Noncanonical Pathways
MSMEG_4570	A0R101	CP_ATPgrasp_1 domain-containing protein	R10, R22, R166	Noncanonical Pathways
MSMEG_4703	A0R1C6	Glycerol-3-phosphate acyltransferase (EC 2.3.1.15)	R67, R173, R754	General Nucleic acid-binding and Metabolism
MSMEG_5075	A0R2D7	ABC transporter ATP-binding protein	R80, R81, R214	General Nucleic acid-binding and Metabolism
MSMEG_5970	A0R4V8	Probable glutamate-1-semialdehyde 2,1-aminomutase	R287, R289, R328	Noncanonical Pathways
RpoB	P60281	DNA-directed RNA polymerase subunit beta'	R273, R508, R529	General Nucleic acid-binding and Metabolism
RpoC	A0QS66	DNA-directed RNA polymerase subunit beta'	R694, R865, R1193	General Nucleic acid-binding and Metabolism
YwnA	A0A0D6HL91	Putative HTH-type transcriptional regulator ywnA	R33, R70, R152	Uncharacterized protein

Table 3: Proteins with greatest total ion score

Gene Name	Protein ID	Protein Name	Ion Score	Phosphosite	Functional Category
MSMEI_3879	I7G417	5-oxoprolinase (ATP-hydrolyzing) (EC 3.5.2.9)	288.7424	R20, R438	Amino acid Metabolism
GlnA	A0R079	Glutamine synthetase (GS)	30.50098	R466	Amino acid Metabolism
AftD	A0QPD4	Alpha-(1->3)-arabinofuranosyltransferase	27.3077	R299, R301, R364	Carbohydrate metabolism
FusA	A4ZHU4	Elongation factor G (EF-G)	26.31937	R373	Translation
RpoC	A0QS66	DNA-directed RNA polymerase subunit beta	26.19321	R694	General Nucleic acid-binding and Metabolism
MSMEG_4757	A0R1H7	Fatty acid synthase	25.0686	R2610	Lipid Biosynthesis
AcnA	A0QX20	Aconitate hydratase A (ACN) (Aconitase)	21.12092	R757	Energy Metabolism
Tuf	A0QS98	Elongation factor Tu (EF-Tu)	19.19775	R291	Translation
recB	A0QS29	RecBCD enzyme subunit RecB	15.4253	R191, R723	General Nucleic acid-binding and Metabolism
Tig	A0R199	Trigger factor (TF) (EC 5.2.1.8) (PPIase)	15.06469	R157	Cell envelope

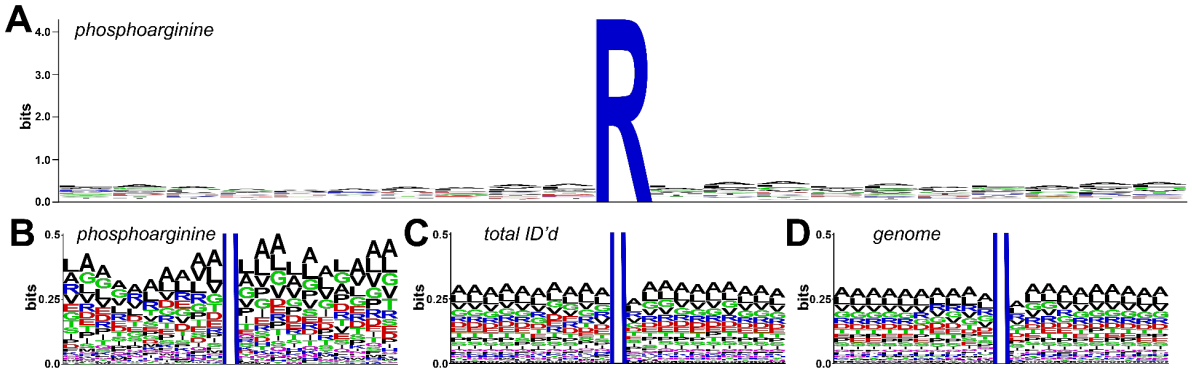


Figure 3. Motif analysis of arginine phosphosites. The amino acid frequency of 21-residue regions surrounding identified phosphoarginine sites is shown as a sequence logo at (**A**) full scale or (**B**) zoomed in to emphasize flaking positions. Comparison datasets show the enrichment of amino acids surrounding (**C**) all Arg residues in proteins identified by MS/MS and (**D**) all Arg residues in the *Msm* proteome.

Arginine Phosphorylation Sites Lack A Consensus Motif – Next, we assessed whether there were recurring sequence motifs surrounding sites of arginine phosphorylation. We aligned 21-residue sequence segments centered on each unique phosphoarginine site and examined the amino acid frequency at each position (45). As shown in **Fig. 3,** no strong enrichment was observed, although Ala, Leu, Val, and Gly were moderately overrepresented (**Fig. 3B**). For comparison, we ran the same analysis on control datasets comprising all arginines in the full set of proteins identified by MS/MS (**Fig. 3C**) or all arginine encoded in the *Msm* genome (**Fig. 3D**). Similar levels of enrichment were observed in all datasets, suggesting that this is an intrinsic feature of the *Msm* proteome and not specific to pArg sequences. Indeed, this pattern likely reflects the fact that Ala, Leu, Val, and Gly are the four most abundant amino acids in the *M. smegmatis* proteome (12%, 10%, 9%, 9% of total, respectively). Thus, amino acid frequency surrounding both pArg and Arg positions is essentially random. These findings are consistent with the variability we observed across phosphoproteomic replicates (**Fig 2C, D**), and suggest that target selection is guided by characteristics other than primary sequence.

Functionally Diverse Proteins Carry Arginine Phosphorylation – To assess whether specific classes of proteins were preferentially modified with phosphoarginine, we performed gene

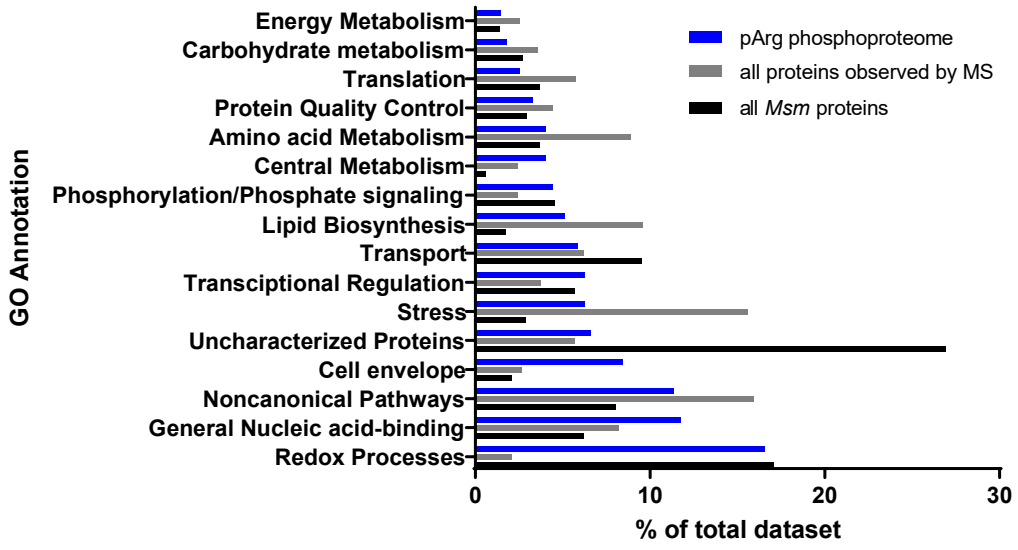


Figure 4: Arginine phosphorylation occurs in diverse targets. Arginine-phosphorylated cellular proteins (blue), all proteins observed by LC-MS/MS (gray) and the full complement of proteins encoded by the *M. smegmatis* genome (black) were subjected to GO annotation and categorized by biological process. The fractional distribution of each dataset among categories is shown.

ontology (GO) analysis on the arginine phosphoproteome using the Blast2GO algorithm (49). As comparison datasets, we performed GO analysis on i) the full set of proteins identified by MS/MS and ii) all proteins encoded by the *Msm* genome. GO terms were assigned based on existing annotations of *Msm* proteins, and/or those of the closest homologs present in the UniProt database.

We found that phosphoarginine modifications occurred on targets with widely diverse cellular functions (**Fig. 4**). Proteins associated with redox reactions were highly represented, including the E1 component of the pyruvate dehydrogenase complex, which was phosphorylated at six positions; catalase peroxidase 1; glycerol-3-phosphate dehydrogenase; and heme chaperone. Nucleic acid-binding proteins were abundant in our dataset as well, including enzymes involved in DNA replication (DNA polymerase I, DNA primase, DNA gyrase subunit A), transcription (RNA polymerase beta subunit RpoC and RpoB), and multiple amino acyl-tRNA ligases (HisS, ProS, AlaS, IleS) (**Fig. 4; Table 1**).

Several functional categories were markedly overrepresented in the pArg dataset compared to their fractional abundance in control datasets, including proteins involved in central metabolism, nucleic-acid binding, and the cell envelope. Interestingly, proteins involved in redox processes were dramatically overrepresented compared to the total set of proteins identified by MS/MS. Uncharacterized proteins were underrepresented, suggesting that some entries in this category are pseudogenes or are otherwise mis-annotated. The apparent enrichment of some functional categories may represent a propensity for these proteins to receive pArg modifications. Alternatively, in cases where categories were more abundant than their occurrence in the total proteome, enrichment may simply reflect high abundance in the cell during normal growth conditions and thus a higher likelihood of observation. Notably, stress response proteins, which were abundant in the *B. subtilis* pArg phosphoproteome (36-38, 69, 70), were not enriched in *M. smegmatis* compared to the total set of identified proteins. Most stress-associated targets were associated with DNA repair, such as Holliday junction ATP-dependent DNA helicase RuvA, UvrD/Rep family helicase, and RecBCD enzyme subunit RecB.

Multiple transmembrane proteins were observed, including single-pass (e.g., penicillin-binding protein PbpB) and multi-pass proteins (eg, ABC transporter MSMEG_5009). In these targets, phosphorylated arginine residues occurred on both cytoplasmic (ESX-1 secretion system protein EccE1, O-antigen export system permease) and periplasmic positions (competence protein MSMEG_4573, ABC-type cobalt transport system MSMEG_6724, ABC transporter MSMEG_5009). It is unlikely that ATP-dependent pArg modifications were installed outside the confines of the plasma membrane. Instead, these proteins may have been phosphorylated prior to membrane insertion or following misfolding/mislocalization within the cytosol.

We also observed arginine phosphorylation of enzymes involved in protein quality control, including Clp protease components. In contrast to its role as a proteolytic marker, there is precedence in *B. subtilis* for pArg to directly modulate protease function: the arginine kinase McsB was found to regulate ClpCP through phosphorylation of the ClpC NTD and D1 ring (37, 40). We

similarly found pArg modification in the *Msm* ClpC1 NTD (Arg93) and D1 ring (Arg365), although these phosphosites are not directly equivalent to those observed in *B. subtilis.* Structures of the virtually identical *Mtb* ClpC1 NTD indicate that Arg93 is solvent-exposed and spatially near the tether to the D1 ring (59). Phosphorylation of Arg93 may modulate interactions with substrates, adaptors, elements of D1, or neighboring NTD modules. Arg365 is conserved in *B. subtilis*, and examination of structures of *B. subtilis* ClpC and *E. coli* ClpA indicate that this position is located near the D1 ATPase active site and is likely solvent exposed. Thus, phosphorylation of Arg365 may modulate ClpC1 ATPase activity and consequently proteolytic activity of ClpC1P1P2. One phosphosite was also found on Arg25 of ClpP1, a solvent-exposed position near the ClpP1 entry portal. No interaction partners are yet known to bind the heptameric face of the ClpP1, so the potential consequences of phosphorylation at this position are unclear.

While a single arginine phosphosite was identified on most targets, several bore multiple modifications. 20 targets bore three or more pArg sites and three proteins carried six sites each (**Table 2**). Multiply-phosphorylated targets included proteins involved in nucleic acid binding and metabolism, noncanonical pathways, and carbohydrate metabolism. Among the proteins with the highest ion scores were hits involved in nucleic acid-binding and metabolism, amino acid metabolism, and translation (**Table 3**). Overall, the number and diversity of pArg-bearing proteins is consistent with the degenerate mode of target selection suggested by motif analysis (**Fig. 3**).

Arginine Phosphorylation Occurs Mostly in Structured Regions - In *B. subtilis*, arginine phosphorylation is upregulated by stress conditions, including heat shock and oxidative stress (37, 38, 63, 69). It has been proposed that pArg modifications in *B. subtilis* primarily occur on damaged or misfolded proteins, directing them for destruction by ClpCP. Whereas Ser, Thr, and Tyr phosphorylation are more likely to occur on unstructured sequences and loops (71, 72), which provide greater accessibility to kinase active sites, the above model predicts that pArg modifications can occur on normally structured sites made accessible by misfolding.

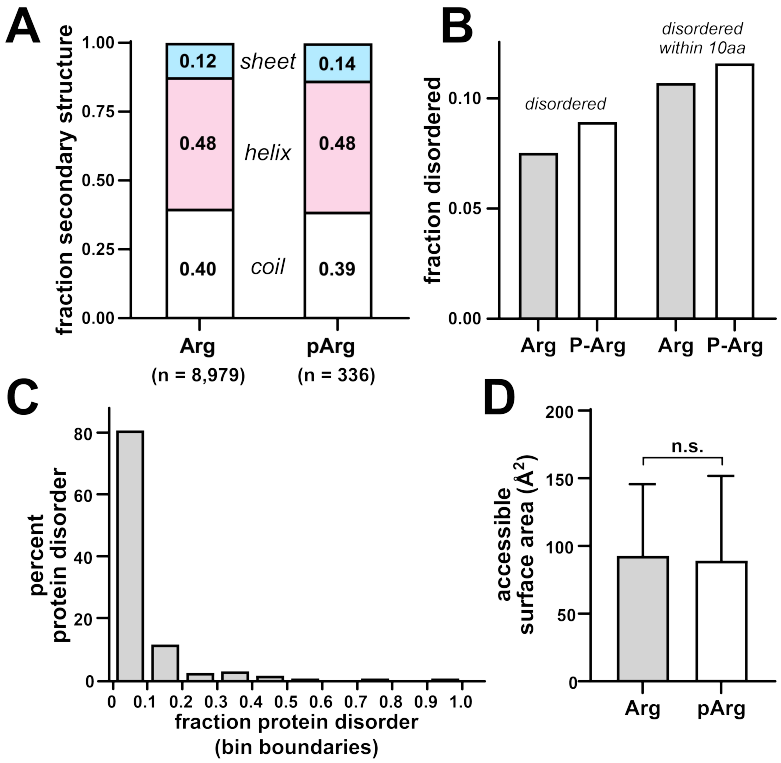


Figure 5. Structural characteristics of phosphoarginine sites. A) Bar segments indicate the fractional distribution of *Msm* phosphoarginine sites or total Arg sites across secondary structure elements predicted by PSIPRED. **B**) The fraction of pArg sites, or total Arg residues in the same set of proteins, located in or within 10 aa of regions predicted to be disordered by the MFDp2 algorithm. **C**) Histogram of predicted fractional sequence disorder in arginine-phosphorylated proteins as predicted by MFDp2. Bin boundaries are labeled. **D**) The average accessible surface area (ASA) of phosphorylated arginines or total arginines among structurally characterized proteins (**Table S2**). For comparison, a fully solvent-exposed Arg has an ASA >200 Å^2. Error bars indicate SD. Statistical significance was assessed by Welch's t-test.

To assess whether *Msm* pArg modifications occur in structured regions, we examined the distribution of pArg sites across secondary structure predicted by PSIPRED (**53**), and compared this to the distribution of all Arg residues present in the same set of proteins. We found that pArg sites occur in strand/helix/coil elements in the same proportions as total Arg residues (**Fig. 5A**), with no bias towards loops, as in *Bacillus subtilis* (37). We next examined the relationship between pArg modifications and predicted disorder. We calculated the fraction of pArg sites and of total Arg residues predicted as disordered by the MFDp2 algorithm (52). A slightly higher percentage of pArg sites occurred in regions of predicted disorder (8.9%), compared to total Arg residues (7.5%) in the same set of proteins, although the overall proportion of sites at such predictions was

low (**Fig. 5B**). We also assessed whether localized phosphoarginines were proximal to disordered regions, and similarly found a slightly higher percentage of pArg sites (11.6%, compared to 10.7% of Arg) within 10 residues of a region with predicted disorder (**Fig. 5B**). Finally, we calculated the overall predicted disorder of each protein sequence (**Fig. 5C**). The majority of proteins bearing pArg modifications (~80%) were predicted to have ≤10% total sequence disorder. These results reveal little propensity for pArg sites to occur on disordered regions of structured protein targets.

Several of the arginine-phosphorylated targets we identified have been structurally characterized (**Supplemental Table S2**). To assess whether pArg modifications occur primarily on solvent-exposed residues, we tabulated the solvent accessibility of 28 pArg sites on 14 structures, and compared these values to the average solvent-accessibility of all structurally resolved arginines in these proteins (**Fig. 5D; Supplemental Table S2**). We found that pArg modifications occurred on residues with a wide range of accessibilities, and there was no statistical difference in the accessibilities of pArg sites compared to Arg residues overall. The observation that at least some pArg modifications occur on residues with low solvent accessibility suggests that these regions must become denatured, or otherwise undergo a large conformational rearrangement, to be accessible to the active site of an arginine kinase. The lack of strong bias towards secondary structure, disorder, or surface accessibility is consistent with a model in which pArg modifications are placed on misfolded, rather than native, protein targets.

Physicochemical Analysis of Arginine Phosphosites – To test whether pArg modifications occurred at sites with particular physiochemical properties, we examined the 10 amino acids upstream and downstream of each localized phosphosite and compared these to the equivalent regions surrounding each of the Arg residues in the full set of identified proteins and in the *Msm* proteome. We calculated the average charge of residues in flanking sequences, and found that the values were close to neutral for all datasets (**Fig. 6B**). Positions flanking pArg sites exhibited a slightly higher average change than Arg sites within identified proteins (+0.02) and slightly lower

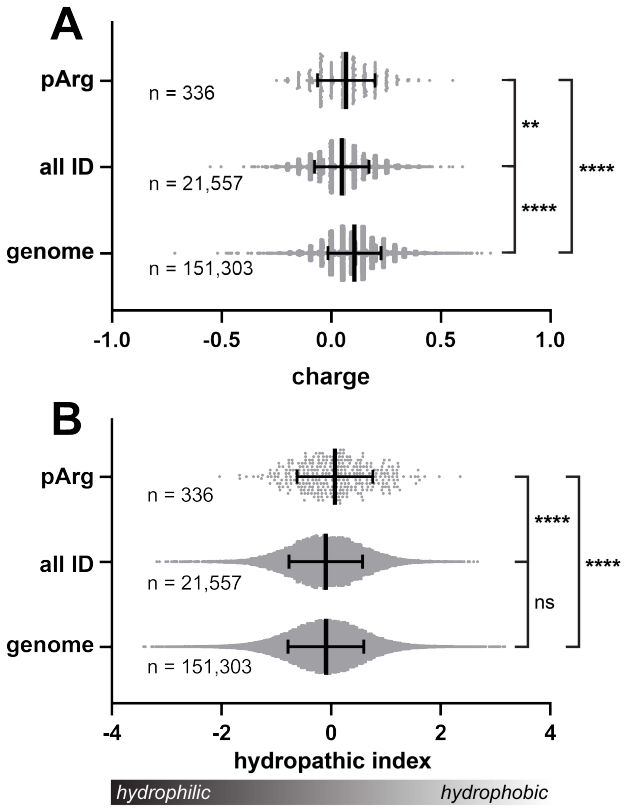


Figure 6. Physiochemical analysis of phosphoarginine sites. A) Violin plots illustrate average charge of the 10 upstream and downstream amino acids flanking observed pArg sites, all Arg residues in proteins identified by MS/MS, or total Arg residues in the *Msm* proteome. Central lines indicate median values, with error bars indicating one standard deviation. **B**) Mean hydrophobicity of residues flanking pArg and control Arg datasets is shown. Significance was assessed by two-tailed Welch's t-test. **** p-value <0.0001; ** <0.01; ns not significant.

value than Arg sites across the entire proteome (-0.03). While these differences were statistically significant, their magnitude is likely too small to serve as a selective criterion for phosphorylation. The differences between observed and theoretical control datasets, in particular, suggests that some aspect of the sample preparation pipeline may introduce bias, such as preferential binding to TiO_2 or increased ionization propensity.

We additionally examined the hydrophobic character of flanking sequences and again observed wide variation, indicating that arginine phosphorylation occurs without strict dependence on the polarity of the phosphosite (**Fig. 6A**). Nevertheless, residues surrounding pArg modifications were, on average, slightly more hydrophobic than control datasets. With this

analysis, there was close agreement between the hydrophobicity of both controls, suggesting less opportunity for bias arising from sample preparation. The observed difference may thus reflect a propensity for phosphorylation on structured elements that become denatured or transiently accessible.

DISCUSSION

Protein phosphorylation is a ubiquitous mechanism of signal propagation and pathway regulation in bacteria (73). While examples of His, Asp, Ser, Thr, and Tyr phosphorylation are widespread, (74-76) the importance of arginine phosphorylation has become apparent only recently. A growing body of data links pArg modifications to protein quality control pathways and spore germination in Firmicutes, including *B. subtilis*.(35-38, 40, 41, 69, 77) However, it is has remained unclear how prevalent pArg modifications are in other bacterial phyla. Here we provide direct evidence for the existence of arginine phosphorylation in *Mycolicibacterium smegmatis*, an actinobacterium.

For arginine phosphorylation to be useful to the cell, it must presumably be applied selectively through the regulated activity of arginine kinases. Many protein kinases recognize substrates through characteristic sequence motifs surrounding the phosphosite. By contrast, our examination of mycobacterial pArg phosphosites revealed no clearly enriched consensus motif. Prior analysis of Arg phosphosites in *B. subtilis* similarly revealed no evident consensus sequence.(37) This degeneracy suggests either that distal interactions guide *Msm* arginine kinases to phosphosites, or that other characteristics of target proteins guide substrate selection. We note that pArg sites were significantly more likely to occur on hydrophobic regions, which may reflect a preference for particular physiochemical properties. However, the magnitude of these differences was small, and the distributions of pArg and control datasets overlap extensively. Thus, physiochemical nature alone is unlikely to enforce strong substrate specificity.

The diversity of proteins observed bearing pArg modifications and their variance among replicate datasets suggests a mode of recognition that is, at least to some degree, independent of protein sequence and identity. Of the 222 *Msm* proteins identified as bearing pArg modifications, only 52 of were identified in ≥2 replicates. These observations are analogous to phosphorylation patterns in *B. subtilis*, where a large number of pArg targets were identified (134 proteins (37) and 87 proteins (40) observed in two independent studies), but with only a 15% overlap between both datasets. The diversity of *Msm* pArg targets extends to functional classification, with GO annotations linking targets to a range of processes and cellular roles. These patterns are consistent with a model in which arginine phosphorylation occurs sporadically on a wide variety of proteins, rather than on a discrete set of targets.

Regardless of the basis for specificity, phosphorylation requires that prospective phosphosites physically dock within a kinase active site. Thus, we expect phosphorylation to be more probable on solvent assessable regions and disordered loops than on buried and structured positions. For example, an analysis of *Mtb* Ser/Thr phosphorylation revealed that phosphosites were 24% more prevalent than non-phosphorylated sites in disordered regions (72, 78). The same study observed a bias toward unstructured coils and loops, with 70% of phosphosites in random coils compared to 56% of non-phosphosites. By contrast, we observed little correlation of pArg sites with secondary structure, predicted disorder, or solvent accessibility. Localized pArg sites were distributed across secondary structure elements in similar proportions as total Arg residues. Only ~9% of pArg sites were found on sites of predicted disorder, a prevalence only slightly higher than Arg overall. Moreover, phosphorylated positions were, on average, no more solvent accessible than other Arg positions, and some positions with low accessibility in published structures were identified as phosphorylated. These structural analyses suggest that many pArg-bearing targets either undergo dramatic conformational rearrangements to expose phosphosites or experience some degree of unfolding prior to modification.

Our data are consistent with a model where arginine phosphorylation is applied to substrates in a predominantly sequence-independent manner to regions that are disordered intrinsically, transiently, or as a consequence of stress. In *B. subtilis* and *S. aureus*, similar sequence independent patterns of phosphorylation were observed, and were upregulated by thermal and oxidative stress (35-38, 40, 41, 69, 77). The levels of pArg modification determined here in unstressed conditions may therefore represent a baseline, and additional experiments will be required to test whether pArg levels are analogously stimulated by stress or other environmental factors in mycobacteria. Similarly, we note that phosphoproteomic studies in Firmicutes utilized deletion strains lacking known arginine phosphatases (^{Sa}PtpB and ^{Bs}YwlE, respectively), which produce higher levels of pArg-bearing proteins. No arginine phosphatases have yet been identified in *Msm* or *Mtb*, thus our approach utilized wild-type *Msm*. In spite of this, we uncovered comparable numbers of pArg modified substrates as prior studies utilizing *B. subtilis* and *S. aureus* deletion strains, suggesting that this modification is at least as prevalent in mycobacteria as in those firmicutes.

What role does arginine phosphorylation play in mycobacterial cells? In *B. subtilis*, pArg tagging is thought to enforce protein quality control by marking misfolded cytosolic proteins for destruction by ClpCP (32-38, 40, 41). Current evidence points to a similar role in mycobacteria. Prior studies demonstrate that the ClpC1P1P2 protease is capable of degrading pArg-bearing protein substrates (41). Given the diverse range of proteins bearing pArg modifications and the degeneracy of the associated phosphosites, it is likely that many of the pArg-bearing targets identified here misfolded spontaneously and were marked for proteolysis. Additionally, the apparent enrichment of functional classifications among pArg-bearing proteins raises the possibility that pArg modifications modulate the function or proteolytic susceptibility of some proteins in their native state, independent of misfolding. For example, proteins associated with central metabolism, nucleic-acid binding, and redox processes are overrepresented in the pArg-bearing dataset. Arginine phosphorylation may regulate these processes. Alternatively, proteins

in these categories may simply be particularly abundant or prone to misfolding. Interestingly, we observed arginine phosphorylation on both ClpC1 and one of its peptidase partners, ClpP1, raising the possibility that arginine kinases both mark proteins for destruction and directly regulate the activity of proteolytic machinery.

This study lays the groundwork for future efforts to expound on the roles of phosphoarginine modifications in mycobacteria. The immediate impediment to further understanding this system is the absence of identifiable orthologs of known arginine kinases or phosphatases. Nevertheless, the diverse target pool identified in this study and our downstream analyses predicts a promiscuous kinase that targets cellular proteins not through of primary sequence or intrinsic residue exposure, but as a result of intermittent or induced unfolding. Future work will be required to decipher how arginine phosphorylation is regulated, and how it contributes to the physiology of mycobacteria and other actinobacterial species.

ACKNOWLEDGEMENTS

We thank members of the Schmitz lab for suggestions and advice; Dr. Papa Nii Asare-Okai for technical help with mass spectrometry; and the Parashar lab for use of the Monolith NT.115. The anti-pArg antibody was a kind gift from Dr. Paul Thompson of the University of Massachusetts Medical School. This work was supported in part by the NIH and NIGMS under Award P20GM104316. The UD Department of Chemistry & Biochemistry Mass Spectrometry core was additionally supported by the NIH and NIGMS under award P30GM110758-02. This content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

TABLE 1: Arginine-phosphorylated proteins in identified in *Msm*

Gene	Protein ID	Protein Description	Phosphosite
		Proteolysis/Protein Quality Control	
clpC1	A0R574	ATP-dependent Clp protease ATP-binding subunit ClpC1	R93, R365
clpP	A0R198	ATP-dependent Clp protease proteolytic subunit	R25
MSMEG_4203	A0QZZ3	Peptidase S8 and S53, subtilisin, kexin, sedolisin	R151, R577
mycP1	A0QNL1	Mycosin-1 (EC 3.4.21.-) (MycP1 protease)	R154
Dop	A0QZ49	Pup deamidase/depupylase (EC 3.4.-.-) (EC 3.5.1.119)	R47, R383
		Amino acid biosynthesis	
glnA	A0R079	Glutamine synthetase (GS)	R466
MSMEG_6197	A0R5H9	Diaminopimelate decarboxylase (EC 4.1.1.20)	R423
purF	A0R4E0	Amidophosphoribosyltransferase (ATase)	R381
ilvD	A0QP06	Dihydroxy-acid dehydratase (DAD) (EC 4.2.1.9)	R311
ilvE	A0R066	Branched-chain-amino-acid aminotransferase	R257
MSMEG_3532	A0QY48	Serine/threonine dehydratase family protein	R22, R34
trpB	A0QX96	Tryptophan synthase beta chain (EC 4.2.1.20)	R172
trpD	A0R048	Anthranilate phosphoribosyltransferase (EC 2.4.2.18)	R209
		Carbohydrate metabolism	
MSMEG_4915	A0R1Y2	Alpha-glucan phosphorylase family protein	R433
aftD	A0QPD4	Alpha-(1->3)-arabinofuranosyltransferase	R299, R301, R364
xylB	A0QXD0	Xylulose kinase (Xylulokinase) (EC 2.7.1.17)	R127, R153, R206, R211
MSMEG_1361	A0QS59	Alpha-mannosidase	R698
pgk	A0QWW3	Phosphoglycerate kinase (EC 2.7.2.3)	R66
		Cell envelope	
MSMEG_4573	A0R104	Competence protein	R116
MSMEG_0249	A0QP26	Integral membrane protein	R41
MSMEG_5818	A0R4F8	Virulence factor Mce family protein	R48
MSMEG_5066	A0R2C8	Integral membrane protein	R165
MSMEG_6140	A0R5C2	Membrane protein	R84
MSMEG_6724	A0R6Z4	ABC-type cobalt transport system	R59
MSMEG_2188	A0QUF5	Integral membrane protein	R274
MSMEG_6216	A0R5J7	Integral membrane protein	R154
MSMEG_1402	A0QS99	Uncharacterized protein	R31, R134
MSMEG_3531	A0QY46	Uncharacterized protein	R196
MSMEG_4218	A0R007	Possible conserved transmembrane protein	R38
MSMEG_3806	A0QYW0	Uncharacterized protein	R135
MSMEG_3570	A0QY85	Uncharacterized protein	R231
MSMEG_6165	A0R5E8	Uncharacterized protein	R152, R162
eccE1	A0QNL0	ESX-1 secretion system protein EccE1	R408
MSMEG_0315	A0QP92	Probable conserved transmembrane protein	R388, R457
MSMEG_6421	A0R646	Transcriptional regulator	R214, R509
MSMEG_5713	A0R457	Major facilitator family protein transporter	R9
MSMEG_6369	A0R5Z3	O-antigen export system, permease protein	R123
PbpB	A0R022	Penicillin-binding protein PbpB (PBP3)	R208
CwsA	A0QNF5	Cell wall synthesis protein CwsA	R59, R115
		Central Metabolism	
MSMEG_3507	A0QY23	Fructose-bisphosphate aldolase class-I (EC 4.1.2.13)	R235
Edd	A0QP89	Phosphogluconate dehydratase (EC 4.2.1.12)	R478
malQ	A0QYI8	4-alpha-glucanotransferase	R320
		Energy Metabolism	
acnA	A0QX20	Aconitate hydratase A (ACN) (Aconitase)	R757
Kgd	A0R2B1	Multifunctional 2-oxoglutarate metabolism enzyme	R618
MSMEG_0848	A0QQR3	NADH ubiquinone oxidoreductase subunit 5	R4, R840, R851
MSMEG_1654	A0QSZ3	Isocitrate dehydrogenase (NADP) (EC 1.1.1.42)	R127, R731
atpG	A0R201	ATP synthase gamma chain	R44
atpC	A0R1Z9	ATP synthase epsilon chain	R105

Lipid Biosynthesis			
fadD28	Q2M5K3	Acyl-CoA synthase (FadD28)	R376
MSMEG_5642	A0R3Y8	Acetyl-coenzyme A carboxylase carboxyl transferase	R288
MSMEG_4254	A0R042	AMP-binding enzyme	R461
MSMEG_2360	A0QUW5	4-coumarate:CoA ligase	R518
MSMEG_2469	A0QV72	AMP-dependent synthetase and ligase	R410
pks5	A0R1E8	Mycocerosic acid synthase-like polyketide synthase	R1357, R1359
MSMEG_4757	A0R1H7	Fatty acid synthase	R273, R362, R1197, R1342, R2610
General Nucleic acid-binding and Metabolism			
polA	A0QYZ2	DNA polymerase I (EC 2.7.7.7)	R321
MSMEG_4912	A0R1X9	Putative helicase	R512
dnaG	O52200	DNA primase (EC 2.7.7.-)	R118, R363
hisS	A0QWK7	Histidine--tRNA ligase (EC 6.1.1.21)	R86, R368
MSMEG_6926	A0R7I2	tRNA adenylyltransferase (EC 2.7.7.72)	R334
Pnp	A0QVQ5	Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8)	R109, R724
gyrA	P48354	DNA gyrase subunit A (EC 5.6.2.2)	R655
rpoC	A0QS66	DNA-directed RNA polymerase subunit beta'	R694, R865, R1193
rpoB	P60281	DNA-directed RNA polymerase subunit beta'	R273, R508, R529
MSMEG_0061	A0QNJ1	Ftsk/spoiiie family protein	R625
proS	A0QVM0	Proline--tRNA ligase (EC 6.1.1.15)	R435
MSMEG_0334	A0QPA9	Acetyl-/propionyl-coenzyme A carboxylase alpha chain	R633
MSMEG_2731	A0QVX6	Uncharacterized protein MSMEG_2731/MSMEI_2664	R291
MSMEG_3578	A0QY93	Putative cyclase	R65
MSMEG_6219	A0R5K0	ATPase family protein associated with various cellular activities	R45
purl	A0R4G4	Phosphoribosylformylglycinamidine synthase subunit PurL	R422
MSMEG_2509	A0QVB0	Mg-chelatase subunits D/I family protein	R99
MSMEG_5075	A0R2D7	ABC transporter ATP-binding protein	R80, R81, R214
MSMEG_4703	A0R1C6	Glycerol-3-phosphate acyltransferase (EC 2.3.1.15)	R67, R173, R754
alaS	A0QWQ4	Alanine--tRNA ligase (EC 6.1.1.7) (Alanyl-tRNA synthetase)	R710, R772, R864
ileS	A0QX46	Isoleucine--tRNA ligase (EC 6.1.1.5) (Isoleucyl-tRNA synthetase)	R430, R652, R769,
MSMEG_5120	A0R2H9	Uncharacterized protein	R263, R286
ERS451418_06140	A0A0D6J306	Putative helicase	R978, R981
D806_012510	A0A2U9PKG0	Type III restriction enzyme, res subunit	R82, R256, R575, R738, R830, R831
ERS451418_05411	A0A0D6IZK9	HNH endonuclease	R242, R310
MSMEG_3926	A0QZ77	Cation-transporting ATPase Pma1 (EC 3.6.3.-)	R56
D806_012420	A0A2U9PKE9	Type III restriction enzyme, res subunit	R461, R535
MSMEG_1383	A0QS81	Endonuclease IV	R146
Phosphorylation/Phosphate signaling			
MSMEG_5849	A0R4I7	RelA_SpoT domain-containing protein	R204, R205
surE	A0R786	Acid phosphatase SurE (EC 3.1.3.2)	R60, R61
MSMEG_4989	A0R252	Sensor histidine kinase	R28
MSMEG_0980	A0QR44	Signal transduction histidine kinase	R242
mtrA	A0QTK2	DNA-binding response regulator MtrA	R147
MSMEG_6257	A0R5N8	Aspartokinase (EC 2.7.2.4)	R284
glpK	A0R5L0	Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase)	R497
MSMEG_5211	A0R2R8	Aminotransferase class-III	R209
Redox Processes			
MSMEG_6557	A0R6I2	Dehydrogenase	R37
MSMEG_1108	A0QRG4	Hydrolase, alpha/beta fold family protein	R28
MSMEG_3977	A0QZC7	Bac_luciferase domain-containing protein	R143

MSMEG_5719	A0R464	Oxidoreductase	R357
MSMEG_2029	A0QU03	3-ketoacyl-ACP/CoA redutase	R10
MSMEG_6134	A0R5B7	NADP-dependent fatty aldehyde dehydrogenase (EC 1.2.1.4)	R141
MSMEG_6491	A0R6B6	Aldehyde dehydrogenase	R348
MSMEG_1603	A0QSU4	Uncharacterized oxidoreductase MSMEG_1603/MSMEI_1564	R250
MSMEG_4891	A0R1V9	Alkyl hydroperoxide reductase C (EC 1.11.1.15)	R116
hemG	A0QW24	Protoporphyrinogen oxidase (EC 1.3.3.4)	R138
MSMEG_2202	A0QUG8	Dimethylaniline monooxygenase (N-oxide-forming) 5	R28
MSMEG_1649	A0QSY8	FAD dependent oxidoreductase	R184
MSMEG_5538	A0R3N8	(NADP+) succinate-semialdehyde dehydrogenase (EC 1.2.1.16)	R428, R442
MSMEG_2080	A0QU53	Putative acyl-CoA dehydrogenase	R201
MSMEG_4149	A0QZU1	Acetaldehyde dehydrogenase 1	R165
MSMEG_4722	A0R1E4	Short-chain dehydrogenase	R251
MSMEG_3805	A0QYW1	Oxidoreductase	R59
MSMEG_3420	A0QXT6	Gluconate 5-dehydrogenase (EC 1.1.1.69)	R114
MSMEG_3106	A0QWY3	Quinone oxidoreductase (EC 1.6.5.5)	R36
MSMEG_5183	A0R2P1	3-Hydroxyacyl-CoA dehydrogenase	R138
MSMEG_6686	A0R6V6	Glutaryl-CoA dehydrogenase (EC 1.3.8.6)	R255, R363
mmsA	A0QV51	Methylmalonate-semialdehyde dehydrogenase (EC 1.2.1.27)	R305
MSMEG_0281	A0QP58	Choline dehydrogenase	R78, R202
MSMEG_2198	A0QUG4	ThiF family protein	R116, R492
MSMEG_4525	A0R0V9	Heme chaperone HemW	R376
MSMEG_5001	A0R263	Pesticide degrading monooxygenase	R351, R375
MSMEG_1795	A0QTC9	2-deoxy-scyllo-inosamine dehydrogenase (EC 1.1.1.-)	R246
MSMEG_0531	A0QPV4	Putative acyl-CoA dehydrogenase	R262, R310
soxA	A0A2U9PT59	Monomeric sarcosine oxidase (EC 1.5.3.1)	R280
D806_067010	A0A2U9Q0Q0	Zinc-binding oxidoreductase	R73
Tmd	A0A0D6GDG8	NADH:flavin oxidoreductase (EC 1.5.8.2)	R320
lgrD_2	A0A0D6G524	Non-ribosomal peptide synthetase	R1154
nemA	A0A0D6IZI2	NADH:flavin oxidoreductase (EC 1.-.-.-)	R209, R279
hisD	A0QTY8	Histidinol dehydrogenase (HDH) (EC 1.1.1.23)	R126
tuaD_2	A0A0D6HF28	UDP-glucose 6-dehydrogenase (EC 1.1.1.22)	R391
MSMEG_1736	A0QT70	Glycerol-3-phosphate dehydrogenase (EC 1.1.5.3)	R174
aceE	A0R0B0	Pyruvate dehydrogenase E1 component (PDH E1 component)	R27, R174, R313, R455, R811, R812
fabG	P71534	3-oxoacyl-(acyl-carrier-protein) reductase FabG I-ACP reductase)	R33
ksdD	A0R4S9	3-oxosteroid 1-dehydrogenase	R281, R340
katG1	A0R609	Catalase-peroxidase 1 (CP 1) (EC 1.11.1.21)	R71
		Stress	
MSMEG_6468	A0R693	Uncharacterized protein	R198
ruvA	A0QWH5	Holliday junction ATP-dependent DNA helicase RuvA	R111, R128
MSMEG_3811	A0QYW6	Universal stress protein family protein, putative	R124
recB	A0QS29	RecBCD enzyme subunit RecB	R191, R723
MSMEG_1941	A0QTR9	Helicase, UvrD/Rep family protein	R76, R391, R578
		Transcriptional Regulation	
MSMEG_0286	A0QP63	GntR-family protein transcriptional regulator	R65
MSMEG_3284	A0QXF7	Transcriptional regulator, MarR family protein	R35
MSMEG_1227	A0QRT1	Transcriptional regulator, GntR family protein	R149
Mfd	A0R3C5	Transcription-repair-coupling factor (TRCF) (EC 3.6.4.-)	R116, R120

MSMEG_0426	A0QPK1	Transcriptional regulator, GntR family protein	R364
MSMEG_0535	A0QPV8	GntR-family protein transcriptional regulator	R172
MSMEG_4405	A0R0J2	RNA polymerase sigma factor	R183
MSMEG_5512	A0R3L1	Magnesium chelatase	R285
MSMEG_1691	A0QT29	Transcriptional regulatory protein	R112, R877
MSMEG_6857	A0R7C3	Putative transcription regulator	R107, R219
MSMEG_5657	A0R402	Transcriptional regulator, LysR family protein	R209, R288
MSMEG_2794	A0QW37	Transcriptional regulator, GntR family protein	R150
ERS451418_03683	A0A0D6IER1	Transcriptional regulatory protein	R219, R496
cueR	A0A0D6IG03	MerR family transcriptional regulator	R165
MSMEG_6236	A0R5L8	Two-component system, regulatory protein	R20
MSMEG_1963	A0QTU1	Putative transcriptional regulatory protein	R40, R165, R850
MSMEG_6244	A0R5M5	TetR-family protein transcriptional regulator	R22
Translation			
fusA	A4ZHU4	Elongation factor G (EF-G)	R373
Tuf	A0QS98	Elongation factor Tu (EF-Tu)	R291
rplJ	A0QS62	50S ribosomal protein L10	R129
Obg	A0R149	GTPase Obg (EC 3.6.5.-) (GTP-binding protein Obg)	R133
rpsE	A0QSG6	30S ribosomal protein S5	R137
rplO	A0QSG8	50S ribosomal protein L15	R60
rpsM	A0QSL5	30S ribosomal protein S13	R94
Transport			
MSMEG_5009	A0R272	ABC transporter	R86, R88
MSMEG_1613	A0QSV3	ABC polar amino acid transporter, inner membrane subunit	R250, R251
MSMEG_1625	A0QSW4	Amino acid transporter, putative	R509
MSMEG_0796	A0QQL3	ABC transporter integral membrane subunit	R18
ssuB	A0R118	Aliphatic sulfonates import ATP-binding protein SsuB (EC 7.6.2.-)	R139
MSMEG_0232	A0QP09	Sugar transporter family protein	R189
MSMEG_2619	A0QVL8	Efflux protein	R280
MSMEG_2570	A0QVG9	Xanthine/uracil permease	R222
MSMEG_4170	A0QZW2	Ribose transport ATP-binding protein RbsA (EC 3.6.3.17)	R476
MSMEG_5660	A0R405	ABC transporter ATP-binding protein	R121
ffh	A0QV32	Signal recognition particle protein (Fifty-four homolog)	R49
MSMEG_4182	A0QZX3	Arabinose-proton symporter	R104, R107
tig	A0R199	Trigger factor (TF) (EC 5.2.1.8) (PPIase)	R40, R157
Miscellaneous Metabolism/ Other Specific Pathways			
MSMEG_3012	A0QWP2	Acetyl-CoA acetyltransferases	R370
MSMEI_0995	I7G667	GCN5-related N-acetyltransferase	R4
MSMEG_2450	A0QV52	Adenosylmethionine--8-amino-7-oxononanoate transaminase	R428
MSMEG_2349	A0QUV4	Glycosyl hydrolase, family protein 57	R239
MSMEG_3175	A0QX52	Pseudouridine synthase (EC 5.4.99.-)	R23
MSMEG_0478	A0QPQ2	Secreted protein, putative	R47, R48
glgB	A0R1Y4	1,4-alpha-glucan branching enzyme GlgB	R322
MSMEG_5970	A0R4V8	Probable glutamate-1-semialdehyde 2,1-aminomutase	R287, R289, R328
MSMEG_1209	A0QRR3	Glycosidase	R420
MSMEG_1050	A0QRA9	SAM-dependent methyltransferase	R247
MSMEG_4256	A0R045	NLP/P60 family protein	R105
MSMEG_4535	A0R0W9	Trehalase (EC 3.2.1.28) (Alpha,alpha-trehalase)	R599
MSMEG_4327	A0R0B4	3-oxoacyl-(acyl-carrier-protein) synthase 1 (EC 2.3.1.41)	R107
MSMEG_6407	A0R632	Cof-like hydrolase	R24
MSMEG_5537	A0R3N6	Conserved integral membrane protein	R185
MSMEG_3505	A0QY21	6-aminohexanoate-cyclic-dimer hydrolase (EC 3.5.2.12)	R57, R341

MSMEG_4510	A0R0U4	Peptide synthetase mbtf	R200, R520, R970, R1009, R1069, R1098
MSMEG_2667	A0QVR7	HpcH/HpaI aldolase/citrate lyase family protein, putative	R153, R280
MSMEG_4570	A0R101	CP_ATPgrasp_1 domain-containing protein	R10, R22, R166
MSMEG_6635	A0R6Q6	Alpha subunit of malonate decarboxylase	R181
MSMEG_3767	A0QYS3	Acyl-CoA synthase	R85, R981
MSMEG_3970	A0QZC2	Glutamyl-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.-)	R181, R223
MSMEG_6392	A0R617	Polyketide synthase	R367, R1586
MSMEI_3879	I7G417	5-oxoprolinase (ATP-hydrolyzing) (EC 3.5.2.9)	R20, R438
MSMEG_2394	A0QUZ6	Putative cystathionine gamma-synthase (EC 2.5.1.48)	R241, R252
hemB	A0QR20	Delta-aminolevulinic acid dehydratase (EC 4.2.1.24)	R19, R62, R63
MSMEG_4168	A0QZW0	Propionate CoA-transferase	R8, R49, R212
glmM_1	A0A0D6GWF0	Phosphomannomutase MrsA (EC 5.4.2.10)	R10
ahcY	A4ZHR8	Adenosylhomocysteinase (EC 3.3.1.1)	R26, R438
MSMEG_1946	A0QTS4	NADH pyrophosphatase (EC 3.6.1.22)	R69
hypF	A0QUN7	Carbamoyltransferase (EC 6.2.-.-)	R554, R758
Uncharacterized proteins			
MSMEG_5490	A0R3J0	Uncharacterized protein	R65
MSMEG_6069	A0R553	CobW/P47K C-domain protein	R186
MSMEG_3133	A0QX11	Uncharacterized protein	R173
MSMEG_4787	A0R1K6	Virulence factor mce family protein	R53
MSMEG_0202	A0QNX9	IS1096, tnpA protein	R109
MSMEG_1419	A0QSB4	Uncharacterized protein	R15
MSMEG_3653	A0QYG7	Uncharacterized protein	R105
MSMEG_0288	A0QP65	FAD dependent oxidoreductase, putative	R257
MSMEG_6596	A0R6M0	Uncharacterized protein	R200
MSMEG_3458	A0QXX4	Uncharacterized protein	R72
MSMEG_4362	A0R0E9	Universal stress protein family protein	R82
MSMEG_0919	Q3I5Q7	HBHA-like protein (Heparin-binding hemagglutinin)	R121
MSMEG_0370	A0QPE5	Uncharacterized protein	R292, R430
MSMEG_1252	A0QRV5	Uncharacterized protein	R62, R157
MSMEG_4660	A0R186	Uncharacterized protein	R118, R146
MSMEG_1 Trentini 1	A0QT19	Endoribonuclease L-PSP superfamily protein	R13
ERS451418_06677	A0A0D6J3W9	Uncharacterized protein	R235
ywnA	A0A0D6HL91	Putative HTH-type transcriptional regulator ywnA	R33, R70, R152

REFERENCES

1. World Health Organization (2020) *Global Tuberculosis Report* (World Health Organization, Geneva)
2. Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger ,T.R., Akerley, B.J., and Sassetti, C.M. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* **7(9),** e1002251.
3. Sassetti, C.M., Boyd, D.H., and Rubin E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular microbiology* **48 (1),** 77-84.
4. DeJesus, M.A., Gerrick, E.R., Xu, W., Park, S.W., Long, J.E. Boutte, C.C., Rubin, E.J., Schnappinger, D., Ehrt, S., Fortune, S.M., Sassetti, C.M., and Ioerger, T.R. (2017) Comprehensive Essentiality Analysis of the *Mycobacterium tuberculosis* Genome via Saturating Transposon Mutagenesis. *mBio* **8(1)**, e02133-16.
5. Raju, R.M., Jedrychowski, M.P., Wei, J.R., Pinkham, J.T., Park, A.S., O'Brien, K., Rehren, G., Schnappinger, D., Gygi S.P., and Rubin, E.J. (2014) Post-translational regulation via Clp protease is critical for survival of *Mycobacterium tuberculosis*. *PLoS Pathog* **10(3)** e1003994.
6. Raju, R. M., Unnikrishnan, M., Rubin, D. H., Krishnamoorthy, V., Kandror, O., Akopian, T. N., Goldberg, A. L., and Rubin, E. J. (2012) *Mycobacterium tuberculosis* ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection. *PLoS pathogens*, **8(2),** e1002511.
7. Famulla, K., Sass, P., Malik, I., Akopian, T., Kandror, O., Alber, M., Hinzen, B., Ruebsamen-Schaeff, H., Kalscheuer, R., Goldberg, A. L., and Brötz-Oesterhelt, H. (2016) Acyldepsipeptide antibiotics kill mycobacteria by preventing the physiological functions of the ClpP1P2 protease. *Molecular microbiology*, **101(2)**, 194–209.
8. Moreira, W., Ngan, G. J., Low, J. L., Poulsen, A., Chia, B. C., Ang, M. J., Yap, A., Fulwood, J., Lakshmanan, U., Lim, J., Khoo, A. Y., Flotow, H., Hill, J., Raju, R. M., Rubin, E. J., and Dick,

T. (2015) Target mechanism-based whole-cell screening identifies bortezomib as an inhibitor of caseinolytic protease in mycobacteria. *mBio*, **6(3**), e00253-15.

9. Gavrish, E., Sit, C. S., Cao, S., Kandror, O., Spoering, A., Peoples, A., Ling, L., Fetterman, A., Hughes, D., Bissell, A., Torrey, H., Akopian, T., Mueller, A., Epstein, S., Goldberg, A., Clardy, J., and Lewis, K. (2014) Lassomycin, a ribosomally synthesized cyclic peptide, kills *Mycobacterium tuberculosis* by targeting the ATP-dependent protease ClpC1P1P2. *Chemistry & biology*, **21(4),** 509–518.
10. Gao, W., Kim, J. Y., Anderson, J. R., Akopian, T., Hong, S., Jin, Y. Y., Kandror, O., Kim, J. W., Lee, I. A., Lee, S. Y., McAlpine, J. B., Mulugeta, S., Sunoqrot, S., Wang, Y., Yang, S. H., Yoon, T. M., Goldberg, A. L., Pauli, G. F., Suh, J. W., Franzblau, S. G., and Cho, S. (2015) The cyclic peptide ecumicin targeting ClpC1 is active against *Mycobacterium tuberculosis* in vivo. *Antimicrobial agents and chemotherapy*, **59(2),** 880–889.
11. Compton, C. L., Schmitz, K. R., Sauer, R. T., and Sello, J. K. (2013) Antibacterial activity of and resistance to small molecule inhibitors of the ClpP peptidase. *ACS chemical biology*, **8(12),** 2669–2677.
12. Moreno-Cinos, C., Goossens, K., Salado, I. G., Van Der Veken, P., De Winter, H., and Augustyns, K. (2019) ClpP Protease, a Promising Antimicrobial Target. *International journal of molecular sciences*, **20(9),** 2232.
13. Sauer, R. T., and Baker, T. A. (2011) AAA+ proteases: ATP-fueled machines of protein destruction. *Annual review of biochemistry*, 80, 587–612.
14. Baker, T. A., and Sauer, R. T. (2012) ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochimica et biophysica acta*, **1823(1),** 15–28.
15. Schmitz, K. R., Carney, D. W., Sello, J. K., and Sauer, R. T. (2014) Crystal structure of *Mycobacterium tuberculosis* ClpP1P2 suggests a model for peptidase activation by AAA+ partner binding and substrate delivery. *Proceedings of the National Academy of Sciences of the United States of America*, **111(43),** E4587–E4595.

16. Akopian, T., Kandror, O., Raju, R. M., Unnikrishnan, M., Rubin, E. J., and Goldberg, A. L. (2012) The active ClpP protease from M. tuberculosis is a complex composed of a heptameric ClpP1 and a ClpP2 ring. *The EMBO journal*, **31(6),** 1529–1541.

17. Schmitz, K. R., and Sauer, R. T. (2014) Substrate delivery by the AAA+ ClpX and ClpC1 unfoldases activates the mycobacterial ClpP1P2 peptidase. *Molecular microbiology*, **93(4),** 617–628.

18. Vahidi, S., Ripstein, Z. A., Juravsky, J. B., Rennella, E., Goldberg, A. L., Mittermaier, A. K., Rubinstein, J. L., and Kay, L. E. (2020) An allosteric switch regulates *Mycobacterium tuberculosis* ClpP1P2 protease function as established by cryo-EM and methyl-TROSY NMR. *Proceedings of the National Academy of Sciences of the United States of America*, **117(11),** 5895–5906.

19. Li, M., Kandror, O., Akopian, T., Dharkar, P., Wlodawer, A., Maurizi, M. R., and Goldberg, A. L. (2016) Structure and Functional Properties of the Active Form of the Proteolytic Complex, ClpP1P2, from *Mycobacterium tuberculosis*. *The Journal of biological chemistry*, **291(14),** 7465–7476.

20. Leodolter, J., Warweg, J., and Weber-Ban, E. (2015) The *Mycobacterium tuberculosis* ClpP1P2 Protease Interacts Asymmetrically with Its ATPase Partners ClpX and ClpC1. *PloS one*, **10(5),** e0125345.

21. Kirstein, J., Hoffmann, A., Lilie, H., Schmidt, R., Rübsamen-Waigmann, H., Brötz-Oesterhelt, H., Mogk, A., and Turgay, K. (2009) The antibiotic ADEP reprogrammes ClpP, switching it from a regulated to an uncontrolled protease. *EMBO molecular medicine*, **1(1),** 37–49.

22. Brötz-Oesterhelt, H., Beyer, D., Kroll, H. P., Endermann, R., Ladel, C., Schroeder, W., Hinzen, B., Raddatz, S., Paulsen, H., Henninger, K., Bandow, J. E., Sahl, H. G., and Labischinski, H. (2005) Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. *Nature medicine*, **11(10),** 1082–1087.

23. Lavey, N. P., Coker, J. A., Ruben, E. A., and Duerfeldt, A. S. (2016) Sclerotiamide: The First Non-Peptide-Based Natural Product Activator of Bacterial Caseinolytic Protease P. *Journal of natural products*, **79(4),** 1193–1197.

24. Akopian, T., Kandror, O., Tsu, C., Lai, J. H., Wu, W., Liu, Y., Zhao, P., Park, A., Wolf, L., Dick, L. R., Rubin, E. J., Bachovchin, W., and Goldberg, A. L. (2015) Cleavage Specificity of *Mycobacterium tuberculosis* ClpP1P2 Protease and Identification of Novel Peptide Substrates and Boronate Inhibitors with Anti-bacterial Activity. *The Journal of biological chemistry*, **290(17),** 11008–11020..

25. Schmitt, E. K., Riwanto, M., Sambandamurthy, V., Roggo, S., Miault, C., Zwingelstein, C., Krastel, P., Noble, C., Beer, D., Rao, S. P., Au, M., Niyomrattanakit, P., Lim, V., Zheng, J., Jeffery, D., Pethe, K., & Camacho, L. R. (2011) The natural product cyclomarin kills *Mycobacterium tuberculosis* by targeting the ClpC1 subunit of the caseinolytic protease. *Angewandte Chemie (International ed. in English)*, **50(26)**, 5889–5891.

26. Wolf, N. M., Lee, H., Choules, M. P., Pauli, G. F., Phansalkar, R., Anderson, J. R., Gao, W., Ren, J., Santarsiero, B. D., Lee, H., Cheng, J., Jin, Y. Y., Ho, N. A., Duc, N. M., Suh, J. W., Abad-Zapatero, C., and Cho, S. (2019) High-Resolution Structure of ClpC1-Rufomycin and Ligand Binding Studies Provide a Framework to Design and Optimize Anti-Tuberculosis Leads. *ACS infectious diseases*, **5(6),** 829–840.

27. Ziemski, M., Leodolter, J., Taylor, G., Kerschenmeyer, A., & Weber-Ban, E. (2021) Genome-wide interaction screen for *Mycobacterium tuberculosis* ClpCP protease reveals toxin-antitoxin systems as a major substrate class. *The FEBS journal*, **288(1),** 111–126

28. Nagpal, J., Paxman, J. J., Zammit, J. E., Alhuwaider, A., Truscott, K. N., Heras, B., and Dougan, D. A. (2019) Molecular and structural insights into an asymmetric proteolytic complex (ClpP1P2) from *Mycobacterium smegmatis*. *Scientific reports*, **9(1),** 18019.

29. Barik, S., Sureka, K., Mukherjee, P., Basu, J., and Kundu, M. (2010) RseA, the SigE specific anti-sigma factor of *Mycobacterium tuberculosis*, is inactivated by phosphorylation-dependent ClpC1P2 proteolysis. *Molecular microbiology*, **75(3),** 592–606.

30. Yee, M., Gopal, P., and Dick, T. (2017) Missense Mutations in the Unfoldase ClpC1 of the Caseinolytic Protease Complex Are Associated with Pyrazinamide Resistance in *Mycobacterium tuberculosis*. *Antimicrobial agents and chemotherapy*, **61(2),** e02342-16.

31. Yamada, Y., and Dick, T. (2017) Mycobacterial Caseinolytic Protease Gene Regulator ClgR Is a Substrate of Caseinolytic Protease. *mSphere*, **2(2),** e00338-16.

32. Gopal, P., Sarathy, J. P., Yee, M., Ragunathan, P., Shin, J., Bhushan, S., Zhu, J., Akopian, T., Kandror, O., Lim, T. K., Gengenbacher, M., Lin, Q., Rubin, E. J., Grüber, G., and Dick, T. (2020) Pyrazinamide triggers degradation of its target aspartate decarboxylase. *Nature communications*, **11(1),** 1661.

33. Lunge, A., Gupta, R., Choudhary, E., and Agarwal, N. (2020) The unfoldase ClpC1 of *Mycobacterium tuberculosis* regulates the expression of a distinct subset of proteins having intrinsically disordered termini. *The Journal of biological chemistry*, **295(28),** 9455–9473.

34. Marsee, J. D., Ridings, A., Yu, T., and Miller, J. M. (2018) *Mycobacterium tuberculosis* ClpC1 N-Terminal Domain Is Dispensable for Adaptor Protein-Dependent Allosteric Regulation. *International journal of molecular sciences*, **19(11),** 3651.

35. Junker, S., Maaβ, S., Otto, A., Michalik, S., Morgenroth, F., Gerth, U., Hecker, M., and Becher, D. (2018) Spectral Library Based Analysis of Arginine Phosphorylations in *Staphylococcus aureus*. *Molecular & cellular proteomics : MCP*, **17(2),** 335–348.

36. Trentini, D. B., Suskiewicz, M. J., Heuck, A., Kurzbauer, R., Deszcz, L., Mechtler, K., and Clausen, T. (2016) Arginine phosphorylation marks proteins for degradation by a Clp protease. *Nature*, **539(7627),** 48–53.

37. Schmidt, A., Trentini, D. B., Spiess, S., Fuhrmann, J., Ammerer, G., Mechtler, K., and Clausen, T. (2014) Quantitative phosphoproteomics reveals the role of protein arginine phosphorylation in the bacterial stress response. *Molecular & cellular proteomics : MCP*, **13(2),** 537–550.

38. Fuhrmann, J., Schmidt, A., Spiess, S., Lehner, A., Turgay, K., Mechtler, K., Charpentier, E., and Clausen, T. (2009) McsB is a protein arginine kinase that phosphorylates and inhibits the heat-shock regulator CtsR. *Science (New York, N.Y.)*, **324(5932),** 1323–1327.

39. Suskiewicz, M. J., Hajdusits, B., Beveridge, R., Heuck, A., Vu, L. D., Kurzbauer, R., Hauer, K., Thoeny, V., Rumpel, K., Mechtler, K., Meinhart, A., and Clausen, T. (2019) Structure of McsB, a protein kinase for regulated arginine phosphorylation. *Nature chemical biology*, **15(5),** 510–518.

40. Elsholz, A. K., Turgay, K., Michalik, S., Hessling, B., Gronau, K., Oertel, D., Mäder, U., Bernhardt, J., Becher, D., Hecker, M., and Gerth, U. (2012) Global impact of protein arginine phosphorylation on the physiology of *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America*, **109(19),** 7451–7456.

41. Weinhäupl, K., Brennich, M., Kazmaier, U., Lelievre, J., Ballell, L., Goldberg, A., Schanda, P., and Fraga, H. (2018) The antibiotic cyclomarin blocks arginine-phosphate-induced millisecond dynamics in the N-terminal domain of ClpC1 from *Mycobacterium tuberculosis*. *The Journal of biological chemistry*, **293(22),** 8379–8393. https://doi.org/10.1074/jbc.RA118.002251.

42. Eddy S. R. (1995) Multiple alignment using hidden Markov models. *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, **3,** 114–120.

43. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, **7,** 539.

44. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*, **25(9),** 1189–1191.

45. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome research*, **14(6),** 1188–1190.

46. Glasgow, J. E., Salit, M. L., and Cochran, J. R. (2016) *In vivo* site-specific protein tagging with diverse amines using an engineered sortase variant. *Journal of the American Chemical Society* **138 (24),** 7496-7499

47. Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nature methods*, **6(5),** 359–362. https://doi.org/10.1038/nmeth.1322

48. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, **5(11),** 976–989.

49. Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)*, **21(18),** 3674–3676.

50. Bairoch, A., and Apweiler, R. (1997) The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucleic acids research*, **25(1),** 31–36.

51. Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., and Yeats, C. (2009) InterPro: the integrative protein signature database. *Nucleic acids research*, **37(Database issue),** D211–D215.

52. Mizianty, M. J., Peng, Z., and Kurgan, L. (2013) MFDp2: Accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. *Intrinsically disordered proteins*, **1(1),** e24428.

53. McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics (Oxford, England)*, **16(4),** 404–405.

54. Mizianty, M. J., Zhang, T., Xue, B., Zhou, Y., Dunker, A. K., Uversky, V. N., and Kurgan, L. (2011) In-silico prediction of disorder content using hybrid sequence representation. *BMC bioinformatics*, **12,** 245.

55. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology*, **372(3),** 774–797.

56. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, **157(1),** 105–132.

57. Moore, D. S. (1986) Amino acid and peptide net charges: A simple calculational procedure. *Biochemical Education* **13 (1)** 10-12.

58. Schlothauer, T., Mogk, A., Dougan, D. A., Bukau, B., and Turgay, K. (2003) MecA, an adaptor protein necessary for ClpC chaperone activity. *Proceedings of the National Academy of Sciences of the United States of America*, **100(5),** 2306–2311.

59. Vasudevan, D., Rao, S. P., and Noble, C. G. (2013) Structural basis of mycobacterial inhibition by cyclomarin A. *The Journal of biological chemistry*, **288(43),** 30883–30891.

60. Bürstner, N., Roggo, S., Ostermann, N., Blank, J., Delmas, C., Freuler, F., Gerhartz, B., Hinniger, A., Hoepfner, D., Liechty, B., Mihalic, M., Murphy, J., Pistorius, D., Rottmann, M., Thomas, J. R., Schirle, M., and Schmitt, E. K. (2015) Gift from Nature: Cyclomarin A Kills Mycobacteria and Malaria Parasites by Distinct Modes of Action. *Chembiochem : a European journal of chemical biology*, **16(17),** 2433–2436.

61. Jung, I. P., Ha, N. R., Kim, A. R., Kim, S. H., and Yoon, M. Y. (2017) Mutation analysis of the interactions between *Mycobacterium tuberculosis* caseinolytic protease C1 (ClpC1) and ecumicin. *International journal of biological macromolecules*, **101,** 348–357.

62. Wolf, N. M., Lee, H., Zagal, D., Nam, J. W., Oh, D. C., Lee, H., Suh, J. W., Pauli, G. F., Cho, S., and Abad-Zapatero, C. (2020) Structure of the N-terminal domain of ClpC1 in complex with the antituberculosis natural product ecumicin reveals unique binding interactions. *Acta crystallographica. Section D, Structural biology*, **76(Pt 5),** 458–471.

63. Fuhrmann, J., Subramanian, V., and Thompson, P. R. (2015) Synthesis and Use of a Phosphonate Amidine to Generate an Anti-Phosphoarginine-Specific Antibody. *Angewandte Chemie (International ed. in English)*, **54(49),** 14715–14718.

64. Potel, C. M., Lin, M. H., Heck, A., and Lemeer, S. (2018) Widespread bacterial protein histidine phosphorylation revealed by mass spectrometry-based proteomics. *Nature methods*, **15(3),** 187–190.

65. Fu, S., Fu, C., Zhou, Q., Lin, R., Han, O., Wang, M., Sun, Y., Liu, Y., and Zhao, Y. (2020) Widespread arginine phosphorylation in human cells—a novel protein PTM revealed by mass spectrometry. *Sci. China Chem.* **63,** 341–346.

66. Thingholm, T. E., Jørgensen, T. J., Jensen, O. N., and Larsen, M. R. (2006) Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nature protocols*, **1(4),** 1929–1935.

67. Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., and Mechtler, K. (2011) Universal and confident phosphorylation site localization using phosphoRS. *Journal of proteome research*, **10(12),** 5354–5362.

68. Bäsell, K., Otto, A., Junker, S., Zühlke, D., Rappen, G. M., Schmidt, S., Hentschker, C., Macek, B., Ohlsen, K., Hecker, M., and Becher, D. (2014) The phosphoproteome and its physiological dynamics in Staphylococcus aureus. *International journal of medical microbiology : IJMM*, **304(2),** 121–132.

69. Fuhrmann, J., Subramanian, V., Kojetin, D. J., and Thompson, P. R. (2016) Activity-Based Profiling Reveals a Regulatory Link between Oxidative Stress and Protein Arginine Phosphorylation. *Cell chemical biology*, **23(8),** 967–977.

70. Schumann W. (2016) Regulation of bacterial heat shock stimulons. *Cell stress & chaperones*, **21(6),** 959–968.

71. Gnad, F., Ren, S., Cox, J., Olsen, J. V., Macek, B., Oroshi, M., and Mann, M. (2007) PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome biology*, **8(11),** R250.

72. Singh, G. P. (2015) Association between intrinsic disorder and serine/threonine phosphorylation in *Mycobacterium tuberculosis*. PeerJ, **3** e724.

73. Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C., and Mijakovic, I. (2019) Protein post-translational modifications in bacteria. *Nature reviews. Microbiology*, **17(11),** 651–664.

74. Chow, K., Ng, D., Stokes, R., and Johnson, P. (1994) Protein tyrosine phosphorylation in Mycobacterium tuberculosis. *FEMS microbiology letters*, **124(2),** 203–207.

75. Canova, M. J., Veyron-Churlet, R., Zanella-Cleon, I., Cohen-Gonsaud, M., Cozzone, A. J., Becchi, M., Kremer, L., and Molle, V. (2008) 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*, **8(3),** 521–533.

76. Nakedi, K. C., Nel, A. J., Garnett, S., Blackburn, J. M., and Soares, N. C. (2015) Comparative Ser/Thr/Tyr phosphoproteomics between two mycobacterial species: the fast growing *Mycobacterium smegmatis* and the slow growing *Mycobacterium bovis* BCG. *Frontiers in microbiology*, **6,** 237.

77. Zhou, B., Semanjski, M., Orlovetskie, N., Bhattacharya, S., Alon, S., Argaman, L., Jarrous, N., Zhang, Y., Macek, B., Sinai, L., and Ben-Yehuda, S. (2019) Arginine dephosphorylation

propels spore germination in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **116(28),** 14228–14237.

78. Prisic, S., Dankwa, S., Schwartz, D., Chou, M. F., Locasale, J. W., Kang, C. M., Bemis, G., Church, G. M., Steen, H., and Husson, R. N. (2010) Extensive phosphorylation with overlapping specificity by Mycobacterium tuberculosis serine/threonine protein kinases. *Proceedings of the National Academy of Sciences of the United States of America*, **107(16),** 7521–7526.