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

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

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

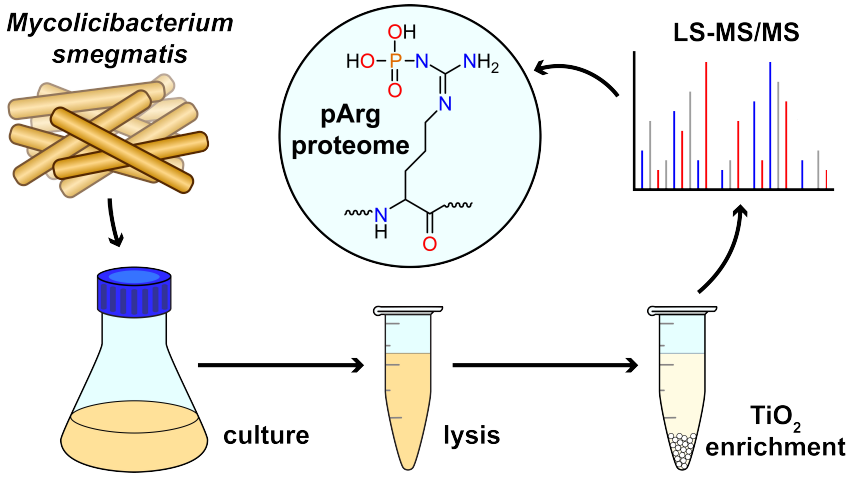
*corresponding author: schmitzk@udel.edu

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

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. While several lines of evidence suggest that ClpC1P1P2 is similarly capable of recognizing phosphoarginine-bearing proteins, 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



65

66 **INTRODUCTION**

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

76 Clp proteases mechanically unfold and destroy native cytosolic proteins (13, 14). These
77 large enzymatic complexes consist of a core peptidase and an associated hexameric ATP-
78 dependent unfoldase. The mycobacterial Clp peptidase, ClpP1P2, is a heteromer composed of
79 distinct ClpP1 and ClpP2 rings that stack face-to-face to create a barrel-shaped tetradecamer (15
80 - 19). ClpC1 is one of two mycobacterial unfoldases (the other is ClpX) that can assemble with
81 ClpP1P2 to form a functional protease. ClpC1 is an 848-residue protein with a globular N-terminal
82 domain (NTD) and two AAA+ (ATPases associated with various cellular activities) modules. Ring-
83 shaped hexamers of ClpC1 dock coaxially on the surface of the ClpP2 face of the peptidase (20).
84 Protein substrates destined for destruction by ClpC1P1P2 must first be recognized by ClpC1, and
85 the unfoldase therefore serves a critical regulatory role in proteolysis.

86 Clp proteases have become major targets of novel drug development against *Mtb* and
87 other pathogenic bacteria. Several classes of antimicrobials specifically target the Clp peptidase,
88 including dysregulators (e.g., acyldepsipeptides (21, 22)), activators (e.g., sclerotiamide (23)),
89 and catalytic inhibitors (e.g., β-lactones and boronate compounds (11, 24)). Importantly, multiple
90 cyclic peptides have been discovered that exhibit anti-*Mtb* activity through dysregulation of ClpC1,

91 including cyclomarin A, lassomycin, ecumicin and rufomycin (9, 10, 25, 26). These compounds
92 bind to the ClpC1 NTD and disrupt proteolysis by uncoupling unfoldase activity from proteolysis,
93 causing uncontrolled degradation of cellular proteins, and/or blocking substrate recognition. The
94 antimicrobial activity of ClpC1-targeting compounds underscores the importance of ClpC1P1P2
95 in mycobacteria. However, the specific proteolytic functions responsible for its essentiality remain
96 obscure, and only a handful of protein substrates have been identified (6, 27-33).

97 An expanded understanding of the physiological roles played by ClpC1P1P2 would bolster
98 efforts to develop antimicrobial compounds. In other bacteria, Clp proteases participate in cellular
99 processes ranging from targeted pathway regulation to cell-wide quality control programs. Multiple
100 mechanisms of substrate recognition have been described, including direct interactions with the
101 unfoldase (6, 31-33) and indirect recognition through proteolytic adaptors (27, 34). Recent studies
102 in *Bacillus subtilis* and related Firmicutes have demonstrated that post-translational arginine
103 phosphorylation marks some proteins for destruction by ClpCP (35-37). Dual phosphoarginine
104 (pArg) binding sites on the *B. subtilis* ClpC NTD allow ClpCP to recognize phosphoarginylated
105 proteins as proteolytic substrates (36). Phosphoarginine-mediated proteolysis in these bacteria is
106 upregulated during stress, through activation of the arginine kinase McsB (38). Phosphoproteomic
107 studies have uncovered widespread arginine phosphorylation during heat stress, implicating
108 ClpCP in turnover of misfolded proteins (36, 37). Additionally, McsB regulates the global stress
109 response through targeted phosphorylation of the negative transcriptional regulators CtsR and
110 HrcA (37-39). Proteolysis of these targets by ClpCP allows transcriptional activation of stress
111 response genes (38). Interestingly, ClpC and ClpP are themselves targets of McsB, and specific
112 pArg sites on ClpC are required for its activation by McsB (37, 40). These studies underline the
113 role of arginine phosphorylation in Firmicutes as both a degradation signal and a regulatory
114 mechanism.

115 Several lines of evidence suggest that an analogous pArg-mediated proteolytic pathway
116 exists in mycobacteria. Sequence and structural data reveal overall homology between the NTDs

117 of *Mtb* ClpC1 and *B. subtilis* ClpC, as well as strong conservation of the residues surrounding the
118 pArg-binding sites (36, 41). Moreover, *in vitro* experiments confirm that the *Mtb* ClpC1 NTD does
119 indeed interact with both free pArg and with arginine-phosphorylated model substrates (41).
120 However, the existence pArg-mediated proteolysis in mycobacteria remains in question, as no
121 mycobacterial McsB homologs are known and phosphoarginine modifications have not yet been
122 described in these bacteria.

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

130

131 **EXPERIMENTAL PROCEDURES**

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

138

139 *Protein purification and binding assays* — A nucleotide sequence encoding the NTD of ClpC1
140 (codons 1 – 147; referred to as $ClpC1^{NTD}$) was amplified from *M. smegmatis* MC^2155 gDNA
141 (ATCC) and cloned into a pET22b-derived vector (EMD Millipore) in frame with a C-terminal
142 LPETGG sortase recognition sequence (46) and 6xHis tag. $ClpC1^{NTD}$ was overexpressed in *E.*

143 *coli* strain ER2566 (NEB) by induction with 0.5 mM isopropyl β-D-1-thiogalactopyranoside at 30°C
144 for 4 h. Cells were harvested by centrifugation, resuspended in lysis buffer (25 mM HEPES, 500
145 mM NaCl, 10 mM imidazole, 10% glycerol, pH 7.5), and lysed by sonication. ClpC1^{NTD} was
146 purified from clarified lysate by Ni-NTA (Marvelgent Biosciences) and anion exchange (Source Q,
147 Cytiva) chromatography. ClpC1^{NTD} was fluorescently labeled via sortase transpeptidation (46), by
148 2 h incubation with sortase A (~1:30 molar ratio) and a 2-fold molar excess of Gly-Gly-Asn-Lys-
149 (fluorescein-isothiocyanate) peptide (Biomatik) in PBS Buffer (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4,
150 137 mM NaCl, 2.7 mM KCl, 10% glycerol, pH 7.4). Excess peptide was removed by size exclusion
151 chromatography (Superdex 75, Cytiva). Purified ClpC1$^{NTD\text{-}FITC}$ was concentrated and stored in
152 CPD Buffer (25 mM HEPES, 200 mM NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, pH 7.0).
153 Binding of phosphoarginine (MilliporeSigma) to 0.1 μM purified ClpC1$^{NTD\text{-}FITC}$ was assayed
154 in CPD (containing 0.05% Tween 20 and 8 mg/ml BSA) by microscale thermophoresis using a
155 Monolith NT.115 (NanoTemper). Thermophoretic data were fit to a quadratic single site binding
156 equation in Prism (Graphpad).
157
158 *Cell Culture Conditions* — Liquid starter cultures of *M. smegmatis* MC^{2}155 were prepared in 10
159 mL Middlebrook broth base (HiMedia) containing 0.2% w/v glucose (TCI), 0.2% v/v glycerol
160 (Fisher Scientific), 0.05% v/v Tween 80, and grown for 60 hours at 37 °C with 250 rpm orbital
161 shaking. Saturated starter cultures were sub-cultured into 200 mL of fresh media (at a starting
162 A_{600} of 0.05), and grown at 37 °C until late log phase or early stationary phase (A_{600} 0.6-1.0). Cells
163 were harvested by centrifugation at 9,000 × g for 20 min at 4 °C, and pellets were resuspended
164 in 5 mL of ice-cold lysis buffer (25 mM HEPES, 200 mM KCl, 10 mM $MgCl_2$, 0.1 mM EDTA, pH
165 7.5) containing 10 mM ATP (Fisher), 200 μL phosphatase inhibitor cocktail (MilliporeSigma), 3
166 mM NaF (Fisher), 20 mM Na_3VO_4 (Acros Organics), 50 μL protease inhibitor cocktail (Thermo
167 Fisher), and 0.2 mM phenylmethylsulfonyl fluoride (PMSF; Acros Organics), and 200 mM
168 dithiotreitol (DTT; Fisher). Cells were lysed at high pressure using a microfluidizer (Microfluidics).

169 Lysates were clarified at 16,000 × g for 30 min at 4 °C, and supernatant was stored at -80 °C prior
170 to further processing. Total protein content was estimated by Bradford assay (Bio-Rad).
171
172 *Filter-Aided Sample Preparation for Mass Spectrometry* — 3 mL replicate samples of clarified cell
173 lysates were prepared for mass spectrometry using filter-aided sample preparation (FASP) (37,
174 47). Reduction of disulfide bonds was achieved by the addition of 10 mM DTT, followed by
175 incubation at 56 °C for 50 min. Afterwards, samples were diluted in 12 mL 0.1 M Tris (Gold
176 Biotechnology), pH 7.5, containing 8 M urea and 55 mM 2-iodoacetamide (both from Acros
177 Organics) in an Amicon filter (10,000 MWCO, MilliporeSigma) and incubated in the dark at room
178 temperature for 45 min to carboxymethylate free cysteines. After alkylation, samples were washed
179 twice with 15 mL 0.1 M Tris, pH 7.5, containing 8 M urea (by 5 min centrifugation at 4,000 × g),
180 then washed twice with 15 mL 0.1 M Tris, pH 7.5, containing 4 M urea, and finally twice with 15
181 mL 100 mM ammonium bicarbonate (Honeywell). After the last wash step, retentate was reduced
182 to 2 mL by spin concentration.
183
184 *In-Solution Trypsin Proteolytic Cleavage* — MS Grade Trypsin Protease (Pierce) was dissolved
185 to obtain a 1 µg/µL stock in 100 mM ammonium bicarbonate. Trypsin digestion of processed *M.*
186 *smegmatis* lysates was performed in an Amicon filter unit at a 100:1 protein:trypsin ratio overnight
187 at 37 °C. Upon digestion, peptides were recovered by centrifugation at 4,000 × g for 10 minutes,
188 followed by the addition of 50 µL of 0.5 M NaCl, and a second round of centrifugation. Eluted
189 samples were dried under vacuum in a SpeedVac centrifuge (Thermo Fisher). To remove residual
190 ammonium bicarbonate and NaCl, samples were desalted using a HyperSep C18 column
191 (Thermo Scientific).
192
193 *TiO_2 Enrichment of Phosphopeptides* — Phosphopeptides were enriched from tryptic peptide
194 samples were using Titansphere TiO_2 beads (5 µm; GL Sciences) in buffer conditions that limit

195 acid hydrolysis of phosphoarginine (36, 37). 5 mg TiO_2 resin was resuspended in 1 mL Binding
196 Buffer (300 mg/mL lactic acid, 12.5% acetic acid, 60% acetonitrile, 0.2% heptafluorobutyric acid,
197 pH 4 with NH_3). Lyophilized peptide samples were re-dissolved in the TiO_2 suspension, incubated
198 for 45 min at 20 °C with gentle agitation, then transferred to graphite spin columns (Thermo
199 Fisher). Unbound peptides were removed by a first wash step with 500 µL Binding Buffer and
200 spun at 1500 × g for 1 min, followed by three wash steps using Wash Solution A (200 mg/ml lactic
201 acid, 75% acetonitrile, 2% trifluoroacetic acid, 2% heptafluorobutyric acid), Wash Solution B (200
202 mg/mL lactic acid, 75% acetonitrile, 10% acetic acid, 0.1% heptafluorobutyric acid, pH 4 with
203 NH_4OH), Wash Solution C (80% acetonitrile, 10% acetic acid), respectively. The resin was then
204 incubated twice with 100 µL Elution Solution (1% NH_4OH, 30 mM ammonium phosphate) for 15
205 minutes, and eluate containing phosphopeptides was collected by centrifugation. Samples were
206 lyophilized and stored at -80 °C prior to mass spectrometry.
207
208 *Liquid chromatography - tandem mass spectrometry (LC-MS/MS)* — Lyophilized samples were
209 resuspended in 0.5% acetic acid, pH 4.5. Samples were then analyzed on a Q Exactive Orbitrap
210 interfaced with Ultimate 3000 Nano-LC system (Thermo Fisher Scientific). 10 µL of trypsin-
211 digested sample was loaded, via an autosampler, onto an Acclaim PepMap RSLC column (75
212 µm x 15 cm nanoViper). Solvents used for the liquid chromatography (LC) were Solvent A (0.1%
213 formic acid in water) and Solvent B (0.1 % formic acid in acetonitrile). Analysis of samples was
214 done using a 150-min gradient running from 2 % Solvent B to 95 % Solvent B at a flow rate of 0.3
215 µL/min. Acquisition of MS data was performed using a data-dependent top10 method, dynamically
216 selecting the most abundant precursor ions from the survey scan for high energy collision-induced
217 dissociation (HCD) fragmentation using a graduated normalized collision energy of 28, 30, and
218 35 eV. Finally, survey scans at a resolution of 70,000 at m/z 200 were obtained on the Q Exactive.
219 LS-MS/MS data was collected in independent biological triplicates.
220

221 *Mass Spectrometry Data Analysis* — Proteomic analysis was performed in the Proteome
222 Discoverer software suite (version 1.4, Thermo Fisher). Raw data was searched against the *M.*
223 *smegmatis* (strain ATCC 700084 / MC^2 155) UniProt Reference Proteome (Proteome ID
224 UP000000757; 6,602 entries in total) using Sequest HT (University of Washington and Thermo
225 Fisher) (48). Iodoacetamide-mediated cysteine carboxymethylation was set as a static
226 modification, phosphorylation of arginine residues was entered as a dynamic modification, and
227 partial trypsinization - with maximum of two missed cleavages - was allowed. Precursor mass
228 tolerance was set at 10 ppm, while allowing fragment ions to have a mass deviation of 0.8 Da for
229 the HCD data. Phosphopeptide hits were re-analyzed using the PhosphoRS node of the
230 Proteome Discoverer software.
231
232 *Experimental Design and Statistical Rationale* — The phosphoarginine discovery workflow was
233 performed on three biological replicates generated from independent *M. smegmatis* cultures. After
234 TiO_2 phosphopeptide enrichment, each replicate was run once on a Q Exactive Orbitrap
235 interfaced with Ultimate 3000 Nano-LC system – yielding one technical replicate per biological
236 sample. Upon re-analysis of phosphopeptide hits, only modifications with a PhosphoRS score
237 ≥75% were selected. Additionally, C-terminal phosphoarginine hits were excluded, and only
238 proteins with positive Sequest HT total ion scores were considered for further analysis. Following
239 these selection criteria, 328 unique phosphoarginine modifications were found on 222 proteins,
240 out of 667 proteins identified by MS/MS (**Table 1; Supplemental Table S1**).
241
242 *Gene Ontology Annotation Analysis* — Gene Ontology (GO) annotation was performed using the
243 OmicsBox software suite (49). Homologs were identified by BLAST against the SwissProt/UniProt
244 database (50), with a minimum expectation value 10^{-3}. Homolog annotations were compiled from
245 the InterPro database (51). BLAST and InterPro results were used to generate GO terms in terms
246 of biological process.

247

248 *Sequence and Physicochemical Analysis of Arginine Phosphosites* — For structural analysis,
249 proteins were processed by the MFDp2 server (52) to evaluate secondary structure (PSIPRED
250 (53), positional disorder (MFDp2), and fractional sequence disorder (Discon (54)). Solvent
251 accessible surface area of arginine residues in published structures was calculated by PISA
252 (**Supplemental Table S2**) (55). For physiochemical analysis, sequences were normalized to a
253 consistent length by incorporating ten residues before and after each arginine, where possible
254 (positions <10 residues from a terminus resulted in shorter sequences; **Supplemental Table S3**).
255 For comparison, control datasets were constructed consisting of i) all arginines present in the 667
256 proteins identified by MS/MS (21,557 total) or ii) all arginines present in the *M. smegmatis*
257 proteome (151,303 total). The average hydrophobicity and the average charge of residues
258 flanking the central arginine at pH 7.0 was calculated using the Kyte-Doolittle scale (56) and
259 published pKa values (57), respectively. Sequence logos of aligned 21-residue segments
260 centered on phosphoarginines and control arginine datasets were generated using WebLogo
261 (45).

262

263 **RESULTS**

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

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

287

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

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

308 *Msm* cell lysates were subjected to a modified filter-aided sample preparation protocol and
309 trypsinized in solution (47). Phosphopeptides were selectively enriched using TiO_2 resin (66).
310 Three biological replicates were analyzed by LC-MS/MS using higher-energy collisional
311 dissociation (HCD) fragmentation. Data analysis was performed using Sequest HT (48), and
312 phosphosite localization was evaluated using the PhosphoRS node of the Proteome Discoverer
313 software suite (67). To reduce the likelihood of incorrect assignment, additional threshold criteria

314 were employed to validate arginine phosphosites: i) peptides with a C-terminal pArg were
315 excluded because manual validation of the corresponding spectra was ambiguous, and because
316 this modification is thought to physiologically impair trypsin digestion (37), ii) only proteins with
317 positive Sequest HT total ion scores were considered for further analysis, and iii) only peptides
318 with a phosphosite localization probability (PhosphoRS scores) ≥75% were selected (68). Using
319 this workflow, we identified 328 unique phosphoarginine sites in 222 of the 667 cellular proteins
320 identified overall (**Fig. 2B; Table 1**; **Supplemental Table S1**). Surprisingly, the list of specific
321 phosphosites varied substantially across replicates, with only three arginine-phosphorylated
322 proteins observed across all three replicates (**Fig. 2C, D**). This variability suggests that arginine
323 phosphorylation occurs sporadically on a large number of proteins, rather than on a discrete set
324 of cellular targets.
325
326 ***Arginine Phosphorylation Sites Lack a Consensus Motif –*** Next, we assessed whether there
327 were recurring sequence motifs surrounding sites of arginine phosphorylation. We aligned 21-
328 residue sequence segments centered on each unique phosphoarginine site and examined the
329 amino acid frequency at each position (45). As shown in **Fig. 3A,** no strong enrichment was
330 observed, although Ala, Leu, Val, and Gly were moderately overrepresented (**Fig. 3B**). For
331 comparison, we ran the same analysis on control datasets comprising all arginines in the full set
332 of proteins identified by MS/MS (**Fig. 3C**) or all arginines encoded in the *Msm* genome (**Fig. 3D**).
333 Similar levels of enrichment were observed in all datasets, suggesting that this is an intrinsic
334 feature of the *Msm* proteome and not specific to pArg sequences. Indeed, this pattern likely
335 reflects the fact that Ala, Leu, Val, and Gly are the four most abundant amino acids in the *M.*
336 *smegmatis* proteome (12%, 10%, 9%, 9% of total, respectively). Thus, amino acid frequency
337 surrounding both pArg and Arg positions is essentially random. These findings are consistent with
338 the variability we observed across phosphoproteomic replicates (**Fig 2C, D**), and suggest that

339 target selection is guided by characteristics other than primary sequence.

340

341 ***Functionally Diverse Proteins Carry Arginine Phosphorylation*** **–** To assess whether specific
342 classes of proteins were preferentially modified with phosphoarginine, we performed gene
343 ontology (GO) analysis on the arginine phosphoproteome using the Blast2GO algorithm (49). As
344 comparison datasets, we performed GO analysis on i) the full set of proteins identified by MS/MS
345 and ii) all proteins encoded by the *Msm* genome. GO terms were assigned based on existing
346 annotations of *Msm* proteins, and/or those of the closest homologs present in the UniProt
347 database.

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

356 Several functional categories were markedly overrepresented in the pArg dataset
357 compared to their fractional abundance in control datasets, including proteins involved in central
358 metabolism, nucleic-acid binding, and the cell envelope. Interestingly, proteins involved in redox
359 processes were dramatically overrepresented compared to the total set of proteins identified by
360 MS/MS. Uncharacterized proteins were underrepresented, suggesting that some entries in this
361 category are pseudogenes or are otherwise mis-annotated. The apparent enrichment of some
362 functional categories may represent a propensity for these proteins to receive pArg modifications.
363 Alternatively, in cases where categories were more abundant than their occurrence in the total
364 proteome, enrichment may simply reflect high abundance in the cell during normal growth

365 conditions and thus a higher likelihood of observation. Notably, stress response proteins, which
366 were abundant in the *B. subtilis* pArg phosphoproteome (36-38, 69, 70), were not enriched in *M.*
367 *smegmatis* compared to the total set of identified proteins. Most stress-associated targets were
368 associated with DNA repair, such as Holliday junction ATP-dependent DNA helicase RuvA,
369 UvrD/Rep family helicase, and RecBCD enzyme subunit RecB.

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

378 We also observed arginine phosphorylation of enzymes involved in protein quality control,
379 including Clp protease components. In contrast to its role as a proteolytic marker, there is
380 precedence in *B. subtilis* for pArg to directly modulate protease function: the arginine kinase McsB
381 was found to regulate ClpCP through phosphorylation of the ClpC NTD and D1 ring (37, 40).
382 Similarly, we found pArg modification in the *Msm* ClpC1 NTD (Arg93) and D1 ring (Arg365),
383 although these phosphosites are not directly equivalent to those observed in *B. subtilis.* Structures
384 of the virtually identical *Mtb* ClpC1 NTD indicate that Arg93 is solvent-exposed and spatially near
385 the tether to the D1 ring (59). Phosphorylation of Arg93 may modulate interactions with
386 substrates, adaptors, elements of D1, or neighboring NTD modules. Arg365 is conserved in *B.*
387 *subtilis*, and examination of structures of *B. subtilis* ClpC and *E. coli* ClpA indicate that this position
388 is located near the D1 ATPase active site and is likely solvent exposed. Thus, phosphorylation of
389 Arg365 may modulate ClpC1 ATPase activity and consequently proteolytic activity of ClpC1P1P2.
390 One phosphosite was also found on Arg25 of ClpP1, a solvent-exposed position near the ClpP1

391 entry portal. No interaction partners are yet known to bind the heptameric face of the ClpP1, so
392 the potential consequences of phosphorylation at this position are unclear.

393 While a single arginine phosphosite was identified on most targets, several bore multiple
394 modifications. 21 targets bore three or more pArg sites, among which one protein bore seven
395 sites and two others carried six sites each (**Table 2**). Multiply-phosphorylated targets included
396 proteins involved in nucleic acid binding and metabolism, noncanonical pathways, and
397 carbohydrate metabolism. Among the proteins bearing sites identified with the highest ion (Xcorr)
398 scores were hits associated with redox processes, nucleic acid binding and metabolism,
399 proteolysis/protein quality control, noncanonical pathways, and the cell envelope (**Table 3**).
400 Overall, the number and diversity of pArg-bearing proteins is consistent with the degenerate mode
401 of target selection suggested by motif analysis (**Fig. 3**).

402

403 ***Arginine Phosphorylation Occurs Mostly in Structured Regions -*** In *B. subtilis*, arginine
404 phosphorylation is upregulated by stress conditions, including heat shock and oxidative stress
405 (37, 38, 63, 69). It has been proposed that pArg modifications in *B. subtilis* primarily occur on
406 damaged or misfolded proteins, directing them for destruction by ClpCP. Whereas Ser, Thr, and
407 Tyr phosphorylation are more likely to occur on unstructured sequences and loops (71, 72), which
408 provide greater accessibility to kinase active sites, the above model predicts that pArg
409 modifications can occur on normally structured sites made accessible by misfolding.

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

417 of pArg sites occurred in regions of predicted disorder (8.9%), compared to total Arg residues
418 (7.5%) in the same set of proteins, although the overall proportion of sites at such predictions was
419 low (**Fig. 5B**). We also assessed whether localized phosphoarginines were proximal to disordered
420 regions, and similarly found a slightly higher percentage of pArg sites (11.6%, compared to 10.7%
421 of Arg) within 10 residues of a region with predicted disorder (**Fig. 5B**). Finally, we calculated the
422 overall predicted disorder of each protein sequence (**Fig. 5C**). The majority of proteins bearing
423 pArg modifications (~80%) were predicted to have ≤10% total sequence disorder. These results
424 reveal little propensity for pArg sites to occur on disordered regions of structured protein targets.
425 Several of the arginine-phosphorylated targets we identified have been structurally
426 characterized (**Supplemental Table S2**). To assess whether pArg modifications occur primarily
427 on solvent-exposed residues, we tabulated the solvent accessibility of 28 pArg sites in 14
428 structures, and compared these values to the average solvent-accessibility of all structurally
429 resolved arginines in these proteins (**Fig. 5D; Supplemental Table S2**). We found that pArg
430 modifications occurred on residues with a wide range of accessibilities, and there was no
431 statistical difference in the accessibilities of pArg sites compared to Arg residues overall. The
432 observation that at least some pArg modifications occur on residues with low solvent accessibility
433 suggests that these regions must become denatured, or otherwise undergo a large
434 conformational rearrangement, to be accessible to the active site of an arginine kinase. The lack
435 of strong bias towards secondary structure, disorder, or surface accessibility is consistent with a
436 model in which pArg modifications are placed on misfolded, rather than native, protein targets.
437

438 ***Physicochemical Analysis of Arginine Phosphosites –*** To test whether pArg modifications
439 occurred at sites with particular physicochemical properties, we examined the 10 amino acids
440 upstream and downstream of each localized phosphosite and compared these to the equivalent
441 regions surrounding each of the Arg residues in the full set of identified proteins and in the *Msm*
442 proteome (**Supplemental Table S3**). We calculated the average charge of residues in flanking

443 sequences, and found that the values were close to neutral for all datasets (**Fig. 6A**). Positions
444 flanking pArg sites exhibited a slightly higher average change than Arg sites within identified
445 proteins (+0.02) and slightly lower value than Arg sites across the entire proteome (-0.03). While
446 these differences were statistically significant, their magnitude is likely too small to serve as a
447 selective criterion for phosphorylation. The differences between observed and theoretical control
448 datasets, in particular, suggests that some aspect of the sample preparation pipeline may
449 introduce bias, such as preferential binding to TiO_2 or increased ionization propensity.

450 We additionally examined the hydrophobic character of flanking sequences and again
451 observed wide variation, indicating that arginine phosphorylation occurs without strict
452 dependence on the polarity of the phosphosite (**Fig. 6B**). Nevertheless, residues surrounding
453 pArg modifications were, on average, slightly more hydrophobic than control datasets. With this
454 analysis, there was close agreement between the hydrophobicity of both controls, suggesting less
455 opportunity for bias arising from sample preparation. The observed difference may thus reflect a
456 propensity for phosphorylation on structured elements that become denatured or transiently
457 accessible.

458

459 **DISCUSSION**

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

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

479 The diversity of proteins observed bearing pArg modifications and their variance among
480 replicate datasets suggests a mode of recognition that is, at least to some degree, independent
481 of protein sequence and identity. Of the 222 *Msm* proteins identified as bearing pArg
482 modifications, only 20 were identified in ≥2 replicates. These observations are analogous to
483 phosphorylation patterns in *B. subtilis*, where a large number of pArg targets were identified (134
484 proteins (37) and 87 proteins (40) observed in two independent studies), but with only a 15%
485 overlap between both datasets. The diversity of *Msm* pArg targets extends to functional
486 classification, with GO annotations linking targets to a range of processes and cellular roles.
487 These patterns are consistent with a model in which arginine phosphorylation occurs sporadically
488 on a wide variety of proteins, rather than on a discrete set of targets.

489 Regardless of the basis for specificity, phosphorylation requires that prospective
490 phosphosites physically dock within a kinase active site. Thus, we expect phosphorylation to be
491 more probable on solvent assessable regions and disordered loops than on buried and structured
492 positions. For example, an analysis of *Mtb* Ser/Thr phosphorylation revealed that phosphosites
493 were 24% more prevalent than non-phosphorylated sites in disordered regions (72, 78). The same

494 study observed a bias toward unstructured coils and loops, with 70% of phosphosites in random
495 coils compared to 56% of non-phosphosites. By contrast, we observed little correlation of pArg
496 sites with secondary structure, predicted disorder, or solvent accessibility. Localized pArg sites
497 were distributed across secondary structure elements in similar proportions as total Arg residues.
498 Only ~9% of pArg sites were found on sites of predicted disorder, a prevalence only slightly higher
499 than Arg overall. Moreover, phosphorylated positions were, on average, no more solvent
500 accessible than other Arg positions, and some positions with low accessibility in published
501 structures were identified as phosphorylated. These structural analyses suggest that many pArg-
502 bearing targets either undergo dramatic conformational rearrangements to expose phosphosites
503 or experience some degree of unfolding prior to modification.

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

518 What role does arginine phosphorylation play in mycobacterial cells? In *B. subtilis*, pArg
519 tagging is thought to enforce protein quality control by marking misfolded cytosolic proteins for

520 destruction by ClpCP (32-38, 40, 41). Current evidence points to a similar role in mycobacteria.
521 Prior studies demonstrate that the ClpC1P1P2 protease is capable of degrading model substrates
522 *in vitro* bearing pArg modifications (41). Given the diverse range of proteins bearing pArg
523 modifications and the degeneracy of the associated phosphosites, it is likely that many of the
524 pArg-bearing targets identified here misfolded spontaneously and were consequently marked for
525 proteolysis. Additionally, the apparent enrichment of functional classes among pArg-bearing
526 proteins raises the possibility that pArg modifications modulate the function or proteolytic
527 susceptibility of some proteins in their native state, independent of misfolding. For example,
528 proteins associated with central metabolism, nucleic-acid binding, and redox processes are
529 overrepresented in the pArg-bearing dataset. Arginine phosphorylation may regulate these
530 processes. Alternatively, proteins in these categories may simply be particularly abundant or
531 prone to misfolding. Interestingly, we observed arginine phosphorylation on both ClpC1 and one
532 of its peptidase partners, ClpP1, raising the possibility that arginine kinases both mark proteins
533 for destruction and directly regulate the activity of proteolytic machinery.

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

543 **ACKNOWLEDGEMENTS**

544 We thank members of the Schmitz lab for advice; VP, AJL, PCB, MP, AD and OU for comments
545 on the manuscript; Dr. Papa Nii Asare-Okai for technical help; and the Parashar lab for use of

546 instrumentation. Sortase A was a kind gift from S Novo and JM Fox. The anti-pArg antibody was
547 a kind gift from Dr. Paul Thompson, UMass Medical School.

548 *Funding and additional information*

549 KRS was supported by NIH NIGMS Award P20GM104316. The UD Department of Chemistry &
550 Biochemistry Mass Spectrometry core was additionally supported by NIH NIGMS award
551 P30GM110758-02. This content is solely the responsibility of the authors and does not
552 necessarily represent the official views of the National Institutes of Health.

553 *Author contributions*

554 E. O. and K. S. designed research; E. O. performed research; E. O. and K. S. analyzed data; E.
555 O. and K. S. wrote the paper.

556

557 **DATA AVAILABILITY**

558 The mass spectrometry data from this publication have been submitted to the ProteomeXchange
559 Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (79),
560 and assigned the identifier PXD025324.

561

562 **SUPPLEMENTAL DATA**

563 This article contains supplemental data.

564

565
566

567 **REFERENCES**

568 1. World Health Organization (2020) *Global Tuberculosis Report* (World Health Organization, Geneva)

569 2. Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger ,T.R., Akerley, B.J., and Sassetti, C.M. (2011) High-
570 resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol
571 catabolism. *PLoS Pathog* **7(9),** e1002251.

572 3. Sassetti, C.M., Boyd, D.H., and Rubin E.J. (2003) Genes required for mycobacterial growth defined by
573 high density mutagenesis. *Molecular microbiology* **48 (1),** 77-84.

574 4. DeJesus, M.A., Gerrick, E.R., Xu, W., Park, S.W., Long, J.E. Boutte, C.C., Rubin, E.J., Schnappinger,
575 D., Ehrt, S., Fortune, S.M., Sassetti, C.M., and Ioerger, T.R. (2017) Comprehensive Essentiality
576 Analysis of the *Mycobacterium tuberculosis* Genome via Saturating Transposon Mutagenesis. *mBio*
577 **8(1)**, e02133-16.

578 5. Raju, R.M., Jedrychowski, M.P., Wei, J.R., Pinkham, J.T., Park, A.S., O'Brien, K., Rehren, G.,
579 Schnappinger, D., Gygi S.P., and Rubin, E.J. (2014) Post-translational regulation via Clp protease
580 is critical for survival of *Mycobacterium tuberculosis*. *PLoS Pathog* **10(3)** e1003994.

581 6. Raju, R. M., Unnikrishnan, M., Rubin, D. H., Krishnamoorthy, V., Kandror, O., Akopian, T. N., Goldberg,
582 A. L., and Rubin, E. J. (2012) *Mycobacterium tuberculosis* ClpP1 and ClpP2 function together in
583 protein degradation and are required for viability in vitro and during infection. *PLoS pathogens*, **8(2),**
584 e1002511.

585 7. Famulla, K., Sass, P., Malik, I., Akopian, T., Kandror, O., Alber, M., Hinzen, B., Ruebsamen-Schaeff, H.,
586 Kalscheuer, R., Goldberg, A. L., and Brötz-Oesterhelt, H. (2016) Acyldepsipeptide antibiotics kill
587 mycobacteria by preventing the physiological functions of the ClpP1P2 protease. *Molecular*
588 *microbiology*, **101(2)**, 194–209.

589 8. Moreira, W., Ngan, G. J., Low, J. L., Poulsen, A., Chia, B. C., Ang, M. J., Yap, A., Fulwood, J.,
590 Lakshmanan, U., Lim, J., Khoo, A. Y., Flotow, H., Hill, J., Raju, R. M., Rubin, E. J., and Dick, T. (2015)
591 Target mechanism-based whole-cell screening identifies bortezomib as an inhibitor of caseinolytic
592 protease in mycobacteria. *mBio*, **6(3),** e00253-15.

593 9. Gavrish, E., Sit, C. S., Cao, S., Kandror, O., Spoering, A., Peoples, A., Ling, L., Fetterman, A., Hughes,
594 D., Bissell, A., Torrey, H., Akopian, T., Mueller, A., Epstein, S., Goldberg, A., Clardy, J., and Lewis,

595 K. (2014) Lassomycin, a ribosomally synthesized cyclic peptide, kills *Mycobacterium tuberculosis* by
596 targeting the ATP-dependent protease ClpC1P1P2. *Chemistry & biology*, **21(4),** 509–518.
597 10. Gao, W., Kim, J. Y., Anderson, J. R., Akopian, T., Hong, S., Jin, Y. Y., Kandror, O., Kim, J. W., Lee, I.
598 A., Lee, S. Y., McAlpine, J. B., Mulugeta, S., Sunoqrot, S., Wang, Y., Yang, S. H., Yoon, T. M.,
599 Goldberg, A. L., Pauli, G. F., Suh, J. W., Franzblau, S. G., and Cho, S. (2015) The cyclic peptide
600 ecumicin targeting ClpC1 is active against *Mycobacterium tuberculosis* in vivo. *Antimicrobial agents*
601 *and chemotherapy*, **59(2),** 880–889.
602 11. Compton, C. L., Schmitz, K. R., Sauer, R. T., and Sello, J. K. (2013) Antibacterial activity of and
603 resistance to small molecule inhibitors of the ClpP peptidase. *ACS chemical biology*, **8(12),** 2669–
604 2677.
605 12. Moreno-Cinos, C., Goossens, K., Salado, I. G., Van Der Veken, P., De Winter, H., and Augustyns, K.
606 (2019) ClpP Protease, a Promising Antimicrobial Target. *International journal of molecular*
607 *sciences*, **20(9),** 2232.
608 13. Sauer, R. T., and Baker, T. A. (2011) AAA+ proteases: ATP-fueled machines of protein
609 destruction. *Annual review of biochemistry*, 80, 587–612.
610 14. Baker, T. A., and Sauer, R. T. (2012) ClpXP, an ATP-powered unfolding and protein-degradation
611 machine. *Biochimica et biophysica acta*, **1823(1),** 15–28.
612 15. Schmitz, K. R., Carney, D. W., Sello, J. K., and Sauer, R. T. (2014) Crystal structure of *Mycobacterium*
613 *tuberculosis* ClpP1P2 suggests a model for peptidase activation by AAA+ partner binding and
614 substrate delivery. *Proceedings of the National Academy of Sciences of the United States of*
615 *America*, **111(43),** E4587–E4595.
616 16. Akopian, T., Kandror, O., Raju, R. M., Unnikrishnan, M., Rubin, E. J., and Goldberg, A. L. (2012) The
617 active ClpP protease from M. tuberculosis is a complex composed of a heptameric ClpP1 and a
618 ClpP2 ring. *The EMBO journal*, **31(6),** 1529–1541.
619 17. Schmitz, K. R., and Sauer, R. T. (2014) Substrate delivery by the AAA+ ClpX and ClpC1 unfoldases
620 activates the mycobacterial ClpP1P2 peptidase. *Molecular microbiology*, **93(4),** 617–628.
621 18. Vahidi, S., Ripstein, Z. A., Juravsky, J. B., Rennella, E., Goldberg, A. L., Mittermaier, A. K., Rubinstein,
622 J. L., and Kay, L. E. (2020) An allosteric switch regulates *Mycobacterium tuberculosis* ClpP1P2

623 protease function as established by cryo-EM and methyl-TROSY NMR. *Proceedings of the National*
624 *Academy of Sciences of the United States of America*, **117(11),** 5895–5906.
625 19. Li, M., Kandror, O., Akopian, T., Dharkar, P., Wlodawer, A., Maurizi, M. R., and Goldberg, A. L. (2016)
626 Structure and Functional Properties of the Active Form of the Proteolytic Complex, ClpP1P2, from
627 *Mycobacterium tuberculosis*. *The Journal of biological chemistry*, **291(14),** 7465–7476.
628 20. Leodolter, J., Warweg, J., and Weber-Ban, E. (2015) The *Mycobacterium tuberculosis* ClpP1P2
629 Protease Interacts Asymmetrically with Its ATPase Partners ClpX and ClpC1. *PloS one*, **10(5),**
630 e0125345.
631 21. Kirstein, J., Hoffmann, A., Lilie, H., Schmidt, R., Rübsamen-Waigmann, H., Brötz-Oesterhelt, H., Mogk,
632 A., and Turgay, K. (2009) The antibiotic ADEP reprogrammes ClpP, switching it from a regulated to
633 an uncontrolled protease. *EMBO molecular medicine*, **1(1),** 37–49.
634 22. Brötz-Oesterhelt, H., Beyer, D., Kroll, H. P., Endermann, R., Ladel, C., Schroeder, W., Hinzen, B.,
635 Raddatz, S., Paulsen, H., Henninger, K., Bandow, J. E., Sahl, H. G., and Labischinski, H. (2005)
636 Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. *Nature*
637 *medicine*, **11(10),** 1082–1087.
638 23. Lavey, N. P., Coker, J. A., Ruben, E. A., and Duerfeldt, A. S. (2016) Sclerotiamide: The First Non-
639 Peptide-Based Natural Product Activator of Bacterial Caseinolytic Protease P. *Journal of natural*
640 *products*, **79(4),** 1193–1197.
641 24. Akopian, T., Kandror, O., Tsu, C., Lai, J. H., Wu, W., Liu, Y., Zhao, P., Park, A., Wolf, L., Dick, L. R.,
642 Rubin, E. J., Bachovchin, W., and Goldberg, A. L. (2015) Cleavage Specificity of *Mycobacterium*
643 *tuberculosis* ClpP1P2 Protease and Identification of Novel Peptide Substrates and Boronate
644 Inhibitors with Anti-bacterial Activity. *The Journal of biological chemistry*, **290(17),** 11008–11020..
645 25. Schmitt, E. K., Riwanto, M., Sambandamurthy, V., Roggo, S., Miault, C., Zwingelstein, C., Krastel, P.,
646 Noble, C., Beer, D., Rao, S. P., Au, M., Niyomrattanakit, P., Lim, V., Zheng, J., Jeffery, D., Pethe, K.,
647 & Camacho, L. R. (2011) The natural product cyclomarin kills *Mycobacterium tuberculosis* by
648 targeting the ClpC1 subunit of the caseinolytic protease. *Angewandte Chemie (International ed. in*
649 *English)*, **50(26)**, 5889–5891.
650 26. Wolf, N. M., Lee, H., Choules, M. P., Pauli, G. F., Phansalkar, R., Anderson, J. R., Gao, W., Ren, J.,
651 Santarsiero, B. D., Lee, H., Cheng, J., Jin, Y. Y., Ho, N. A., Duc, N. M., Suh, J. W., Abad-Zapatero,

652 C., and Cho, S. (2019) High-Resolution Structure of ClpC1-Rufomycin and Ligand Binding Studies
653 Provide a Framework to Design and Optimize Anti-Tuberculosis Leads. *ACS infectious*
654 *diseases*, **5(6),** 829–840.
655 27. Ziemski, M., Leodolter, J., Taylor, G., Kerschenmeyer, A., & Weber-Ban, E. (2021) Genome-wide
656 interaction screen for *Mycobacterium tuberculosis* ClpCP protease reveals toxin-antitoxin systems as
657 a major substrate class. *The FEBS journal*, **288(1),** 111–126
658 28. Nagpal, J., Paxman, J. J., Zammit, J. E., Alhuwaider, A., Truscott, K. N., Heras, B., and Dougan, D. A.
659 (2019) Molecular and structural insights into an asymmetric proteolytic
660 complex (ClpP1P2) from *Mycobacterium smegmatis*. *Scientific reports*, **9(1),** 18019.
661 29. Barik, S., Sureka, K., Mukherjee, P., Basu, J., and Kundu, M. (2010) RseA, the SigE specific anti-sigma
662 factor of *Mycobacterium tuberculosis*, is inactivated by phosphorylation-dependent ClpC1P2
663 proteolysis. *Molecular microbiology*, **75(3),** 592–606.
664 30. Yee, M., Gopal, P., and Dick, T. (2017) Missense Mutations in the Unfoldase ClpC1 of the Caseinolytic
665 Protease Complex Are Associated with Pyrazinamide Resistance in *Mycobacterium*
666 *tuberculosis*. *Antimicrobial agents and chemotherapy*, **61(2),** e02342-16.
667 31. Yamada, Y., and Dick, T. (2017) Mycobacterial Caseinolytic Protease Gene Regulator ClgR Is a
668 Substrate of Caseinolytic Protease. *mSphere*, **2(2),** e00338-16.
669 32. Gopal, P., Sarathy, J. P., Yee, M., Ragunathan, P., Shin, J., Bhushan, S., Zhu, J., Akopian, T., Kandror,
670 O., Lim, T. K., Gengenbacher, M., Lin, Q., Rubin, E. J., Grüber, G., and Dick, T. (2020) Pyrazinamide
671 triggers degradation of its target aspartate decarboxylase. *Nature communications*, **11(1),** 1661.
672 33. Lunge, A., Gupta, R., Choudhary, E., and Agarwal, N. (2020) The unfoldase ClpC1 of *Mycobacterium*
673 *tuberculosis* regulates the expression of a distinct subset of proteins having intrinsically disordered
674 termini. *The Journal of biological chemistry*, **295(28),** 9455–9473.
675 34. Marsee, J. D., Ridings, A., Yu, T., and Miller, J. M. (2018) *Mycobacterium tuberculosis* ClpC1 N-
676 Terminal Domain Is Dispensable for Adaptor Protein-Dependent Allosteric Regulation. *International*
677 *journal of molecular sciences*, **19(11),** 3651.
678 35. Junker, S., Maaβ, S., Otto, A., Michalik, S., Morgenroth, F., Gerth, U., Hecker, M., and Becher, D. (2018)
679 Spectral Library Based Analysis of Arginine Phosphorylations in *Staphylococcus aureus*. *Molecular*
680 *& cellular proteomics : MCP*, **17(2),** 335–348.

681 36. Trentini, D. B., Suskiewicz, M. J., Heuck, A., Kurzbauer, R., Deszcz, L., Mechtler, K., and Clausen, T.
682 (2016) Arginine phosphorylation marks proteins for degradation by a Clp
683 protease. *Nature*, **539(7627),** 48–53.
684 37. Schmidt, A., Trentini, D. B., Spiess, S., Fuhrmann, J., Ammerer, G., Mechtler, K., and Clausen, T.
685 (2014) Quantitative phosphoproteomics reveals the role of protein arginine phosphorylation in the
686 bacterial stress response. *Molecular & cellular proteomics : MCP*, **13(2),** 537–550.
687 38. Fuhrmann, J., Schmidt, A., Spiess, S., Lehner, A., Turgay, K., Mechtler, K., Charpentier, E., and
688 Clausen, T. (2009) McsB is a protein arginine kinase that phosphorylates and inhibits the heat-shock
689 regulator CtsR. *Science (New York, N.Y.)*, **324(5932),** 1323–1327.
690 39. Suskiewicz, M. J., Hajdusits, B., Beveridge, R., Heuck, A., Vu, L. D., Kurzbauer, R., Hauer, K., Thoeny,
691 V., Rumpel, K., Mechtler, K., Meinhart, A., and Clausen, T. (2019) Structure of McsB, a protein kinase
692 for regulated arginine phosphorylation. *Nature chemical biology*, **15(5),** 510–518.
693 40. Elsholz, A. K., Turgay, K., Michalik, S., Hessling, B., Gronau, K., Oertel, D., Mäder, U., Bernhardt, J.,
694 Becher, D., Hecker, M., and Gerth, U. (2012) Global impact of protein arginine phosphorylation on
695 the physiology of *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United*
696 *States of America*, **109(19),** 7451–7456.
697 41. Weinhäupl, K., Brennich, M., Kazmaier, U., Lelievre, J., Ballell, L., Goldberg, A., Schanda, P., and
698 Fraga, H. (2018) The antibiotic cyclomarin blocks arginine-phosphate-induced millisecond dynamics
699 in the N-terminal domain of ClpC1 from *Mycobacterium tuberculosis*. *The Journal of biological*
700 *chemistry*, **293(22),** 8379–8393. https://doi.org/10.1074/jbc.RA118.002251.
701 42. Eddy S. R. (1995) Multiple alignment using hidden Markov models. *Proceedings. International*
702 *Conference on Intelligent Systems for Molecular Biology*, **3,** 114–120.
703 43. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert,
704 M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality
705 protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, **7,** 539.
706 44. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version
707 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford,*
708 *England)*, **25(9),** 1189–1191.

709 45. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo
710 generator. *Genome research*, **14(6),** 1188–1190.
711 46. Glasgow, J. E., Salit, M. L., and Cochran, J. R. (2016) *In vivo* site-specific protein tagging with diverse
712 amines using an engineered sortase variant. *Journal of the American Chemical Society* **138 (24),**
713 7496-7499
714 47. Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation
715 method for proteome analysis. *Nature methods*, **6(5),** 359–362. https://doi.org/10.1038/nmeth.1322
716 48. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral
717 data of peptides with amino acid sequences in a protein database. *Journal of the American Society*
718 *for Mass Spectrometry*, **5(11),** 976–989.
719 49. Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005) Blast2GO: a
720 universal tool for annotation, visualization and analysis in functional genomics
721 research. *Bioinformatics (Oxford, England)*, **21(18),** 3674–3676.
722 50. Bairoch, A., and Apweiler, R. (1997) The SWISS-PROT protein sequence data bank and its supplement
723 TrEMBL. *Nucleic acids research*, **25(1),** 31–36.
724 51. Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U.,
725 Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud,
726 A., Letunic, I., Lonsdale, D., Lopez, R., and Yeats, C. (2009) InterPro: the integrative protein signature
727 database. *Nucleic acids research*, **37(Database issue),** D211–D215.
728 52. Mizianty, M. J., Peng, Z., and Kurgan, L. (2013) MFDp2: Accurate predictor of disorder in proteins by
729 fusion of disorder probabilities, content and profiles. *Intrinsically disordered proteins*, **1(1),** e24428.
730 53. McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) The PSIPRED protein structure prediction
731 server. *Bioinformatics (Oxford, England)*, **16(4),** 404–405.
732 54. Mizianty, M. J., Zhang, T., Xue, B., Zhou, Y., Dunker, A. K., Uversky, V. N., and Kurgan, L. (2011) In-
733 silico prediction of disorder content using hybrid sequence representation. *BMC bioinformatics*, **12,**
734 245.
735 55. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline
736 state. *Journal of molecular biology*, **372(3),** 774–797.

737 56. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a
738 protein. *Journal of molecular biology*, **157(1),** 105–132.
739 57. Moore, D. S. (1986) Amino acid and peptide net charges: A simple calculational procedure. *Biochemical*
740 *Education* **13 (1)** 10-12.
741 58. Schlothauer, T., Mogk, A., Dougan, D. A., Bukau, B., and Turgay, K. (2003) MecA, an adaptor protein
742 necessary for ClpC chaperone activity. *Proceedings of the National Academy of Sciences of the*
743 *United States of America*, **100(5),** 2306–2311.
744 59. Vasudevan, D., Rao, S. P., and Noble, C. G. (2013) Structural basis of mycobacterial inhibition by
745 cyclomarin A. *The Journal of biological chemistry*, **288(43),** 30883–30891.
746 60. Bürstner, N., Roggo, S., Ostermann, N., Blank, J., Delmas, C., Freuler, F., Gerhartz, B., Hinniger, A.,
747 Hoepfner, D., Liechty, B., Mihalic, M., Murphy, J., Pistorius, D., Rottmann, M., Thomas, J. R., Schirle,
748 M., and Schmitt, E. K. (2015) Gift from Nature: Cyclomarin A Kills Mycobacteria and Malaria Parasites
749 by Distinct Modes of Action. *Chembiochem : a European journal of chemical biology*, **16(17),** 2433–
750 2436.
751 61. Jung, I. P., Ha, N. R., Kim, A. R., Kim, S. H., and Yoon, M. Y. (2017) Mutation analysis of the interactions
752 between *Mycobacterium tuberculosis* caseinolytic protease C1 (ClpC1) and ecumicin. *International*
753 *journal of biological macromolecules*, **101,** 348–357.
754 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
755 Abad-Zapatero, C. (2020) Structure of the N-terminal domain of ClpC1 in complex with the
756 antituberculosis natural product ecumicin reveals unique binding interactions. *Acta crystallographica.*
757 *Section D, Structural biology*, **76(Pt 5),** 458–471.
758 63. Fuhrmann, J., Subramanian, V., and Thompson, P. R. (2015) Synthesis and Use of a Phosphonate
759 Amidine to Generate an Anti-Phosphoarginine-Specific Antibody. *Angewandte Chemie (International*
760 *ed. in English)*, **54(49),** 14715–14718.
761 64. Potel, C. M., Lin, M. H., Heck, A., and Lemeer, S. (2018) Widespread bacterial protein histidine
762 phosphorylation revealed by mass spectrometry-based proteomics. *Nature methods*, **15(3),** 187–
763 190.

764 65. Fu, S., Fu, C., Zhou, Q., Lin, R., Han, O., Wang, M., Sun, Y., Liu, Y., and Zhao, Y. (2020) Widespread
765 arginine phosphorylation in human cells—a novel protein PTM revealed by mass spectrometry. *Sci.*
766 *China Chem.* **63,** 341–346.
767 66. Thingholm, T. E., Jørgensen, T. J., Jensen, O. N., and Larsen, M. R. (2006) Highly selective enrichment
768 of phosphorylated peptides using titanium dioxide. *Nature protocols*, **1(4),** 1929–1935.
769 67. Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., and Mechtler, K. (2011)
770 Universal and confident phosphorylation site localization using phosphoRS. *Journal of proteome*
771 *research*, **10(12),** 5354–5362.
772 68. Bäsell, K., Otto, A., Junker, S., Zühlke, D., Rappen, G. M., Schmidt, S., Hentschker, C., Macek, B.,
773 Ohlsen, K., Hecker, M., and Becher, D. (2014) The phosphoproteome and its physiological dynamics
774 in Staphylococcus aureus. *International journal of medical microbiology : IJMM*, **304(2),** 121–132.
775 69. Fuhrmann, J., Subramanian, V., Kojetin, D. J., and Thompson, P. R. (2016) Activity-Based Profiling
776 Reveals a Regulatory Link between Oxidative Stress and Protein Arginine Phosphorylation. *Cell*
777 *chemical biology*, **23(8),** 967–977.
778 70. Schumann W. (2016) Regulation of bacterial heat shock stimulons. *Cell stress & chaperones*, **21(6),**
779 959–968.
780 71. Gnad, F., Ren, S., Cox, J., Olsen, J. V., Macek, B., Oroshi, M., and Mann, M. (2007) PHOSIDA
781 (phosphorylation site database): management, structural and evolutionary investigation, and
782 prediction of phosphosites. *Genome biology*, **8(11),** R250.
783 72. Singh, G. P. (2015) Association between intrinsic disorder and serine/threonine phosphorylation in
784 *Mycobacterium tuberculosis*. PeerJ, **3** e724.
785 73. Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C., and Mijakovic, I. (2019)
786 Protein post-translational modifications in bacteria. *Nature reviews. Microbiology*, **17(11),** 651–664.
787 74. Chow, K., Ng, D., Stokes, R., and Johnson, P. (1994) Protein tyrosine phosphorylation in
788 Mycobacterium tuberculosis. *FEMS microbiology letters*, **124(2),** 203–207.
789 75. Canova, M. J., Veyron-Churlet, R., Zanella-Cleon, I., Cohen-Gonsaud, M., Cozzone, A. J., Becchi, M.,
790 Kremer, L., and Molle, V. (2008) The *Mycobacterium tuberculosis* serine/threonine kinase PknL
791 phosphorylates Rv2175c: mass spectrometric profiling of the activation loop phosphorylation sites
792 and their role in the recruitment of Rv2175c. *Proteomics*, **8(3),** 521–533.

793 76. Nakedi, K. C., Nel, A. J., Garnett, S., Blackburn, J. M., and Soares, N. C. (2015) Comparative
794 Ser/Thr/Tyr phosphoproteomics between two mycobacterial species: the fast growing
795 *Mycobacterium smegmatis* and the slow growing *Mycobacterium bovis* BCG. *Frontiers in*
796 *microbiology*, **6,** 237.
797 77. Zhou, B., Semanjski, M., Orlovetskie, N., Bhattacharya, S., Alon, S., Argaman, L., Jarrous, N., Zhang,
798 Y., Macek, B., Sinai, L., and Ben-Yehuda, S. (2019) Arginine dephosphorylation propels spore
799 germination in bacteria. *Proceedings of the National Academy of Sciences of the United States of*
800 *America*, **116(28),** 14228–14237.
801 78. Prisic, S., Dankwa, S., Schwartz, D., Chou, M. F., Locasale, J. W., Kang, C. M., Bemis, G., Church, G.
802 M., Steen, H., and Husson, R. N. (2010) Extensive phosphorylation with overlapping specificity by
803 Mycobacterium tuberculosis serine/threonine protein kinases. *Proceedings of the National Academy*
804 *of Sciences of the United States of America*, **107(16),** 7521–7526.
805 79. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti,
806 A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T.,
807 Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A. F., and Vizcaíno, J. A. (2019)
808 The PRIDE database and related tools and resources in 2019: improving support for quantification
809 data. *Nucleic acids research*, **47(D1),** D442–D450.
810

811 TABLES

812

Table 1. Arginine-phosphorylated proteins identified in *Msm*

Gene	Protein ID	Protein Description	Phosphosite(s)
		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, R170, 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 reductase	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.-)	R170, 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
		Noncanonical 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

815

Table 2: Proteins with 3 or more phosphoarginine modifications				
Gene Name	**Protein ID**	**Protein Name**	**Phosphosite**	**Functional Category**
D806_012510	A0A2U9PKG0	Type III restriction enzyme, res subunit	R82, R170, R256, R575, R738, R830, R831	General Nucleic acid-binding and Metabolism
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
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

816

817

Table 3: Phosphoarginine sites with the greatest ion (Xcorr) scores

Protein ID	Protein Name	Localized Phosphopeptide Sequence	Phosphosite(s)	Ion (Xcorr) Score	Functional Category
A0A0D6GDG8	NADH:flavin oxidoreductase	EGVLDIIGAArPSIADPFLPAK	R320	2.45	Redox processes
A0R104	Competence protein	LMFrAALTELDGAPVSGrVMV	R116, R130	2.42	Cell envelope
A0QZU1	Acetaldehyde dehydrogenase 1	SAGPGTrANIDEFTETTALALESVG	R165	2.22	General Nucleic acid-binding and Metabolism
A0QZC7	Bac_luciferase domain-containing protein	LLSDPDHNPTPrQLPPPILVAGA	R143	2.20	Redox processes
A0QX46	Isoleucine--tRNA ligase	SrFWDEDPDAIDTLHTVLEVTcrL	R769	2.18	General Nucleic acid-binding and Metabolism
A0QYW1	Oxidoreductase, short chain dehydrogenase/reductase family protein	rPELVDVPGAHVLRLDITDQWS	R59	2.17	Redox processes
A0A0D6IZK9	HNH endonuclease	TRLDDLILIcANcHrMIHR	R310	2.16	General Nucleic acid-binding and Metabolism
A0QR20	Delta-aminolevulinic acid dehydratase	DSLrrAAADAVAAGVGGLMLFGVPL	R62, R63	2.14	Noncanonical Pathways
A0QWP2	Acetyl-CoA acetyltransferases	IMTTLVNALHQrGAR	R370	2.13	Noncanonical Pathways
A0QZ49	Pup deamidase/depupylase	ENLTWQAPrLHLVDLQYSDVRL	R383	2.11	Proteolysis/Protein Quality Control

FIGURES

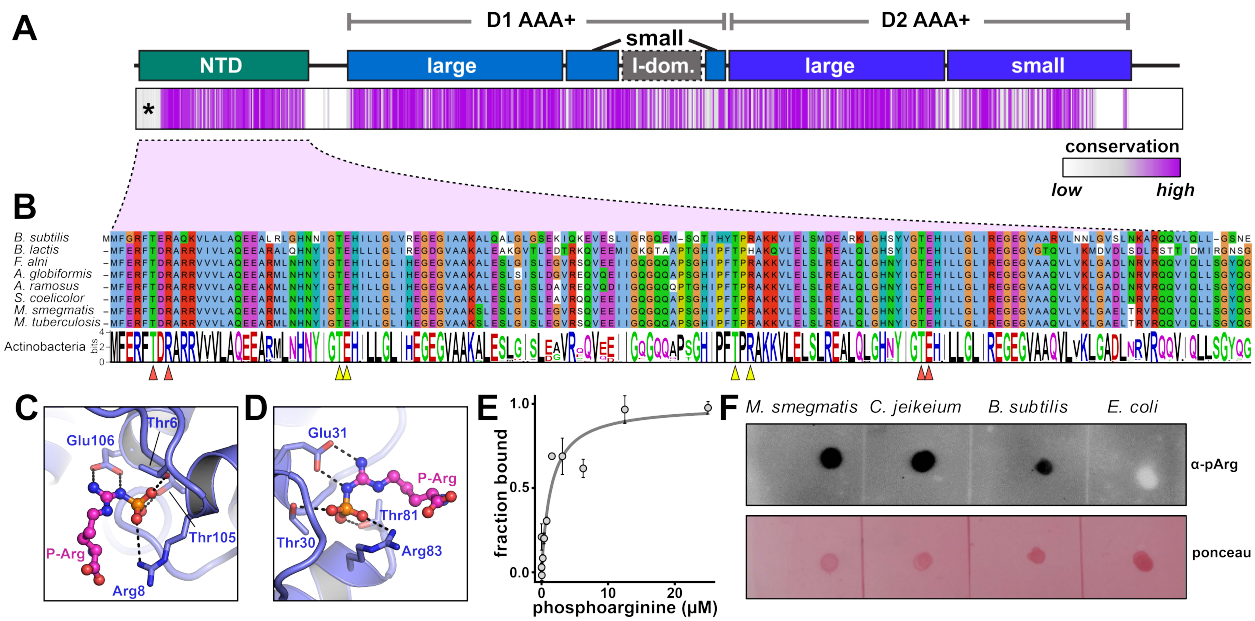
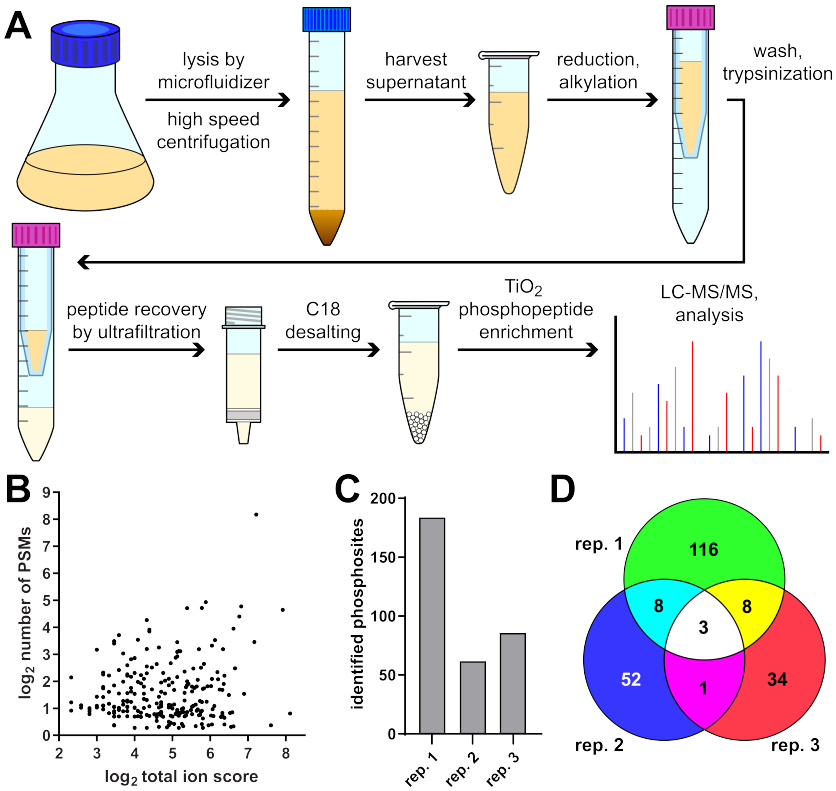


Figure 1. **Indirect evidence for phosphoarginine in actinobacteria**. **A**) The domain organization of ClpC1 is shown above a plot of amino acid conservation among actinobacterial ClpC1 orthologs, where conservation at each alignment position is plotted as a purple vertical strip. The apparent poor conservation at the beginning of the NTD (marked by *) likely reflects mis-annotation of the start site in some entries. **B**) The sequence of the *Bacillus subtilis* ClpC NTD 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 NTD (5HBN) were modeled on the *M. tuberculosis* ClpC1 NTD (6PBQ) putative 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.

837



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

843

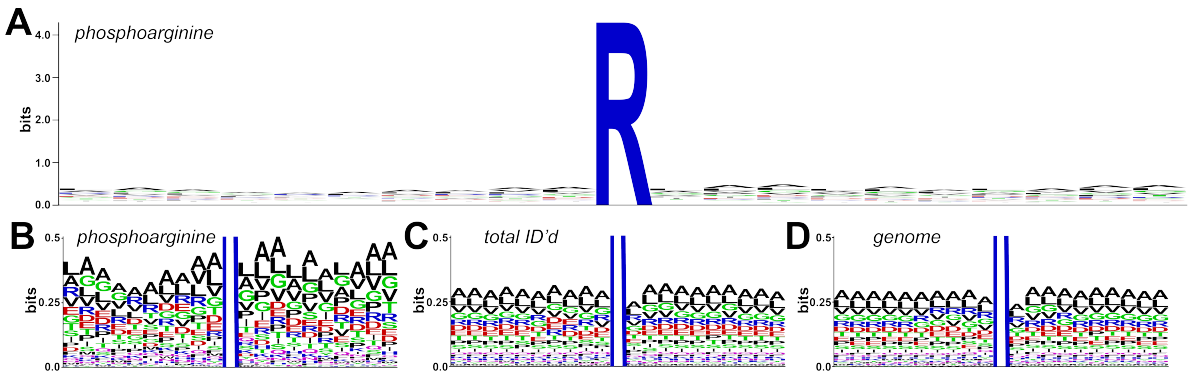


Figure 3. Motif analysis of arginine phosphosites. The amino acid frequency of 21-residue regions surrounding identified phosphoarginine sites is shown as a sequence logo at (**A**) full scale or (**B**) zoomed in to emphasize flanking positions. Comparison datasets show the enrichment of amino acids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

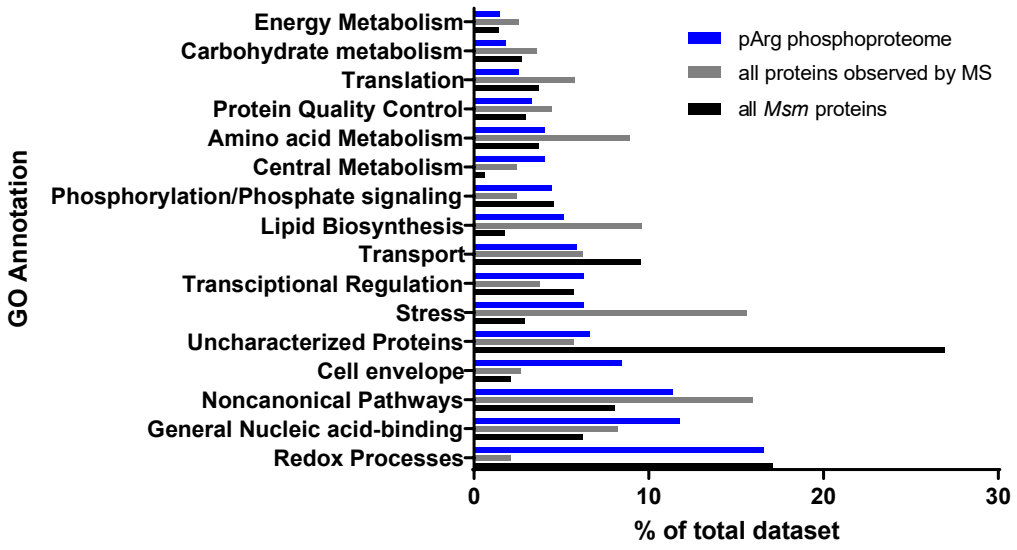
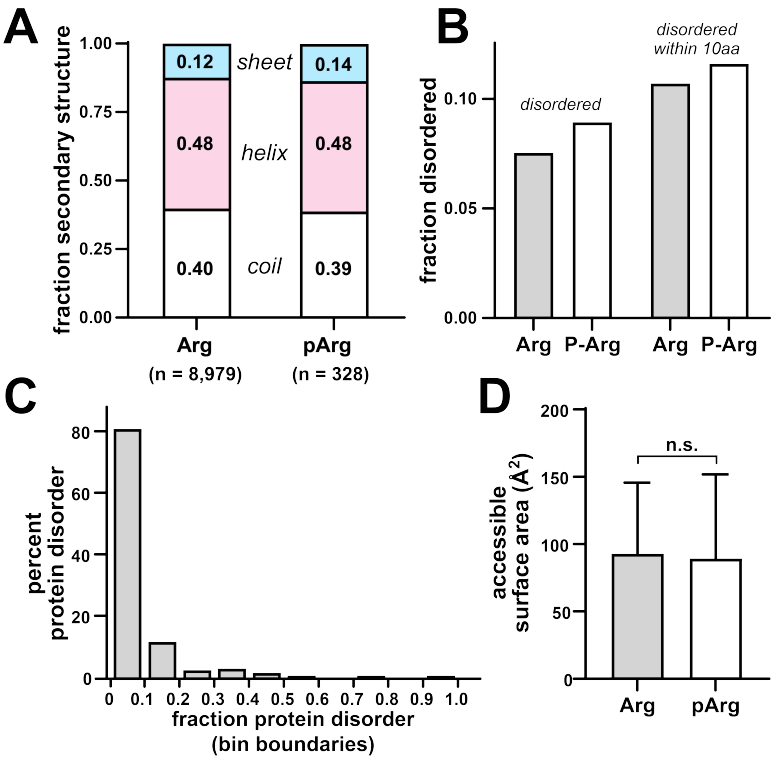


Figure 4: Arginine phosphorylation occurs in diverse targets. Arginine-phosphorylated cellular proteins (blue), all proteins observed by LC-MS/MS (gray) and the full complement of proteins encoded by the *M. smegmatis* genome (black) were subjected to GO annotation and categorized by biological process.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

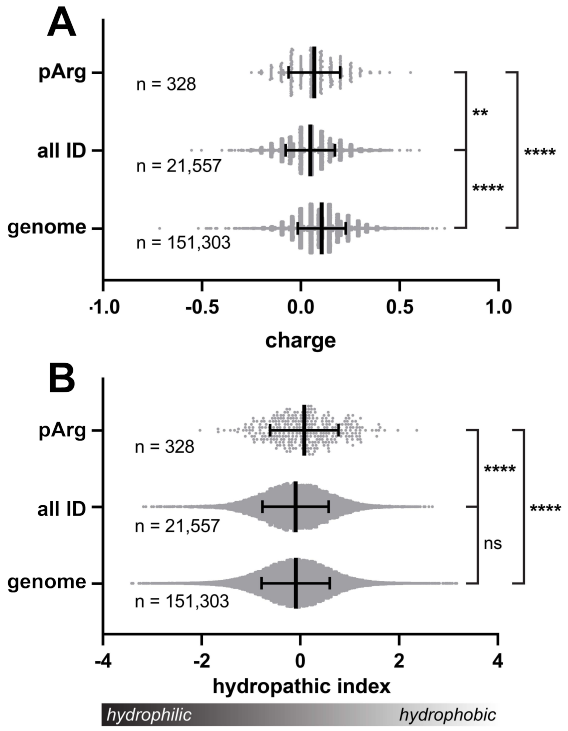
868



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

877

878



879

880 **Figure 6. Physiochemical analysis of phosphoarginine sites. A**) Violin plots illustrate average charge
881 of the 10 upstream and downstream amino acids flanking observed pArg sites, all Arg residues in proteins
882 identified by MS/MS, or total Arg residues in the *Msm* proteome. Central lines indicate median values, with
883 error bars indicating one standard deviation. **B**) Mean hydrophobicity of residues flanking pArg and control
884 Arg datasets is shown. Significance was assessed by two-tailed Welch's t-test. **** p-value <0.0001; **
885 <0.01; ns not significant.

886

887