# Identification of arginine phosphorylation in *Mycolicibacterium smegmatis*

Emmanuel C. Ogbonna[1], Henry R. Anderson[2] 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 several functionally diverse targets within the *Msm* proteome. Interestingly, phosphoarginine modifications are not upregulated by heat stress, suggesting divergent roles in mycobacteria and *Bacillus*. Our findings provide new evidence supporting the existence of phosphoarginine-mediated proteolysis by ClpC1P1P2 in mycobacteria and other actinobacterial species.

65 **Introduction**

66 Tuberculosis is a leading cause of worldwide infectious mortality, ranking above HIV/AIDS
67 and Ebola, and is one of the top ten leading causes of death overall (WHO, 2021). Advances in
68 diagnosis, vaccinations and therapeutics have reduced tuberculosis morbidity and mortality.
69 However, the prevalence of multidrug-resistance in the causative bacterium, *Mycobacterium*
70 *tuberculosis* (*Mtb*), remains high (WHO, 2021). These statistics underscore the urgent need to
71 discover new drugs and exploit new molecular targets. One promising target is the mycobacterial
72 Clp protease. Several studies show that Clp protease components are strictly essential for *Mtb*
73 viability (Sassetti et al., 2003; Griffin et al., 2011; Raju et al., 2012; Raju et al., 2014; DeJesus et
74 al., 2017) and are viable targets for anti-*Mtb* therapeutics (Compton et al., 2013; Gavrish et al,
75 2014; Gao et al, 2015; Moreira et al., 2015; Famulla et al., 2016; Moreno-Cinos et al., 2019).

76 Clp proteases mechanically unfold and destroy native cytosolic proteins (Sauer and Baker;
77 2011; Baker and Sauer 2012). These large enzymatic complexes consist of a core peptidase and
78 an associated hexameric ATP-dependent unfoldase. The mycobacterial Clp peptidase, ClpP1P2,
79 is a heteromer composed of distinct ClpP1 and ClpP2 rings that stack face-to-face to create a
80 barrel-shaped tetradecamer (Akopian et al., 2012; Schmitz et al., 2014; Schmitz et al., 2014b; Li
81 et al., 2016; Vahidi et al., 2020). ClpC1 is one of two mycobacterial unfoldases (the other is ClpX)
82 that can assemble with ClpP1P2 to form a functional protease. ClpC1 is an 848-residue protein
83 with a globular N-terminal domain (NTD) and two AAA+ (ATPases associated with various cellular
84 activities) modules. Ring-shaped hexamers of ClpC1 dock coaxially on the surface of the ClpP2
85 face of the peptidase (Leodolter et al., 2015; Nagpal et al., 2019). Protein substrates destined for
86 destruction by ClpC1P1P2 must first be recognized by ClpC1. The unfoldase therefore serves a
87 critical regulatory role in proteolysis.

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

90 including dysregulators (e.g., acyldepsipeptides (Brötz-Oesterhelt et al., 2005; Kirstein et al,
91 2009)), activators (e.g., sclerotiamide (Lavey et al., 2016)), and catalytic inhibitors (e.g., β-
92 lactones and boronate compounds (Compton et al., 2013; Akopian et al., 2015)). Importantly,
93 multiple cyclic peptides have been discovered that exhibit anti-*Mtb* activity through dysregulation
94 of ClpC1, including cyclomarin A, lassomycin, ecumicin, rufomycin and metamarin (Schmitt et al.,
95 2011; Gavrish et al., 2014; Gao et al., 2015; Wolf et al., 2019; Li et al., 2021). These compounds
96 bind to the ClpC1 NTD and disrupt proteolysis by uncoupling unfoldase activity from proteolysis
97 or by causing uncontrolled degradation of cellular proteins. The antimicrobial activity of ClpC1-
98 targeting compounds underscores the importance of ClpC1P1P2 in mycobacteria. However, the
99 specific proteolytic functions responsible for its essentiality remain obscure, and only a handful of
100 protein substrates have been identified (Barik et al., 2010; Gopal et al., 2020; Lunge et al., 2020;
101 Ziemski et al., 2021).

102 An expanded understanding of the physiological roles played by ClpC1P1P2 would bolster
103 efforts to develop antimicrobial compounds. In other bacteria, Clp proteases participate in cellular
104 processes ranging from targeted pathway regulation to cell-wide protein quality control (Gerth et
105 al., 2008; Trentini et al., 2016). Multiple mechanisms of substrate recognition have been
106 described, including direct interactions with the unfoldase (Gopal et al., 2020; Lunge et al., 2020)
107 and indirect recognition with aid of adaptors (Marsee et al., 2018; Ziemski et al., 2021). Recent
108 studies in *Bacillus subtilis* and related Firmicutes have demonstrated that post-translational
109 arginine phosphorylation marks some proteins for destruction by ClpCP (Schmidt et al., 2014;
110 Trentini et al., 2016; Junker et al., 2018). Dual phosphoarginine (pArg) binding sites on the *B.*
111 *subtilis* ClpC NTD allow ClpCP to recognize phosphoarginylated proteins as proteolytic substrates
112 (Trentini et al., 2016). Phosphoarginine-mediated proteolysis in these bacteria is upregulated
113 during stress through activation of the arginine kinase McsB (Fuhrmann et al., 2009).
114 Phosphoproteomic studies have uncovered widespread arginine phosphorylation during heat
115 stress, implicating ClpCP in turnover of misfolded proteins (Schmidt et al., 2014; Trentini et al.,

116 2016; Trentini et al., 2018). Additionally, McsB regulates the global stress response through
117 targeted phosphorylation of the negative transcriptional regulators CtsR and HrcA (Fuhrmann et
118 al., 2009; Schmidt et al., 2014; Suskiewicz et al., 2019). Proteolysis of these targets by ClpCP
119 allows transcriptional activation of stress response genes (Fuhrmann et al., 2009). Interestingly,
120 ClpC and ClpP are themselves targets of McsB, and specific pArg sites on ClpC are required for
121 its activation by McsB (Schmidt et al., 2014, Elsholz et al., 2012). These studies underline the role
122 of arginine phosphorylation in Firmicutes as both a degradation signal and a regulatory
123 mechanism.

124 Several lines of evidence suggest that an analogous pArg-mediated proteolytic pathway
125 exists in mycobacteria. Sequence and structural data reveal overall homology between the NTDs
126 of *Mtb* ClpC1 and *B. subtilis* ClpC, as well as strong conservation of the residues surrounding the
127 pArg-binding sites (Trentini et al., 2016; Weinhäupl et al., 2018). Moreover, *in vitro* experiments
128 confirm that the *Mtb* ClpC1 NTD does indeed interact with both free pArg and with arginine-
129 phosphorylated model substrates (Weinhäupl et al., 2018). However, the existence of pArg-
130 mediated proteolysis in mycobacteria remains in question, as no mycobacterial McsB homologs
131 are known and phosphoarginine modifications have not yet been described in these bacteria.

132 Here, we confirm the existence of post-translational phosphoarginine modifications in
133 *Mycolicibacterium smegmatis* (*Msm*), a nonpathogenic surrogate of *Mtb.* Using a phosphopeptide
134 enrichment workflow coupled to shotgun phosphoproteomics, we identify arginine phosphosites
135 on several protein targets within the *Msm* proteome. Our findings suggest that these modifications
136 are widespread among actinobacterial species.

137

138 **Materials and Methods**

139 *ClpC1 Sequence Analysis* — Actinobacterial homologs of *Msm* ClpC1 were identified using
140 HMMER (Eddy, 1995) and aligned using Clustal Omega (Sievers et al., 2011). Fragmentary
141 sequences were omitted from analysis. To reduce overrepresentation of similar taxa (e.g.,
142 multiple *Mtb* strains), alignments were pruned such that no two sequences had greater than 90%
143 sequence identity. Positional conservation scores were exported from Jalview (Waterhouse et al.,
144 2009) and plotted as a heatmap in Prism (Graphpad). The sequence logo of the NTD region was
145 generated using WebLogo (Crooks et al., 2004).

146 *Protein Purification and Binding Assays* — The nucleotide sequence encoding the NTD of ClpC1
147 (codons 1 – 147; referred to as $ClpC1^{NTD}$) was amplified from *M. smegmatis* $MC^2155$ gDNA
148 (ATCC) and cloned into a pET22b-derived vector (EMD Millipore) in frame with a C-terminal
149 LPETGG sortase recognition sequence (Glasgow et al., 2016) and 6xHis tag. $ClpC1^{NTD}$ was
150 overexpressed in *E. coli* strain ER2566 (NEB) by induction with 0.5 mM isopropyl β-D-1-
151 thiogalactopyranoside at 30°C for 4 h. Cells were harvested by centrifugation, resuspended in
152 Lysis Buffer (25 mM HEPES, 500 mM NaCl, 10 mM imidazole, 10% glycerol, pH 7.5), and lysed
153 by sonication. $ClpC1^{NTD}$ was purified from clarified lysate by Ni-NTA (Marvelgent Biosciences)
154 and anion exchange (Source Q, Cytiva) chromatography. $ClpC1^{NTD}$ was fluorescently labeled via
155 sortase transpeptidation (Glasgow et al., 2016), by 2 h incubation with sortase A (~1:30 molar
156 ratio) and a 2-fold molar excess of a Gly-Gly-Asn-Lys-(fluorescein-isothiocyanate) peptide
157 (Biomatik) in PBS Buffer (10 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$, 137 mM NaCl, 2.7 mM KCl, 10%
158 glycerol, pH 7.4). Excess peptide was removed by size exclusion chromatography (Superdex 75,
159 Cytiva). Purified $ClpC1^{NTD\text{-}FITC}$ was concentrated and stored in CPD Buffer (25 mM HEPES, 200
160 mM NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, pH 7.0). Binding of phosphoarginine (MilliporeSigma) to
161 0.1 μM purified $ClpC1^{NTD\text{-}FITC}$ was assayed in CPD supplemented with 0.05% Tween 20 and 8
162 mg/ml BSA by microscale thermophoresis using a Monolith NT.115 (NanoTemper) (Jerabek-

163 Willemsen et al., 2014). Thermophoretic data were fit to a quadratic single site binding equation
164 in Prism (Graphpad).

165 *Cell Culture Conditions* — Liquid starter cultures of *M. smegmatis* $MC^2155$ were prepared in 20
166 mL Middlebrook broth base (HiMedia) containing 0.2% v/v glycerol (Fisher Scientific), 0.05% v/v
167 Tween 80, and grown for 48 hours at 37 °C with 250 rpm orbital shaking. Saturated starter cultures
168 were sub-cultured into 900 mL of fresh media at a starting $A_{600}$ ~0.02, and grown at 37 °C until
169 $A_{600}$ ~1.0. 500 mL of culture was collected and added to 400 mL of fresh media. Unstressed
170 control samples were further grown at 37 °C, while samples for heat stress were grown at 50 °C
171 for 21 hours. Cells were harvested by centrifugation at 9,000 × g for 20 min at 4 °C, and pellets
172 were resuspended in 5 mL of ice-cold lysis buffer (25 mM HEPES, 200 mM KCl, 10 mM $MgCl_2$,
173 0.1 mM EDTA, pH 7.5) containing 10 mM ATP (Fisher), 200 µL phosphatase inhibitor cocktail
174 (MilliporeSigma), 30 mM sodium pervanadate (Acros Organics), 3 mM NaF and 3 mM sodium
175 pyrophosphate (both from Fisher). Cells were lysed at high pressure using a microfluidizer
176 (Microfluidics). Lysates were clarified at 15,000 × g for 30 min at 4 °C, and supernatant was stored
177 at -80 °C prior to further processing. Total protein content was estimated by Bradford assay (Bio-
178 Rad).

179 *Filter-Aided Sample Preparation for Mass Spectrometry* — Replicate samples of clarified cell
180 lysates (approximately 10 mg total protein) were prepared for mass spectrometry using filter-aided
181 sample preparation (FASP) (Wiśniewski et al., 2009; Schmidt et al., 2014). Reduction of disulfide
182 bonds was achieved by the addition of 200 mM dithiothreitol (DTT; Fisher), followed by incubation
183 at 56 °C for 50 min. Afterwards, samples were diluted in 7 mL of 0.1 M Tris (Gold Biotechnology),
184 pH 7.5 containing 8 M urea and 25 mM 2-iodoacetamide (both from Acros Organics) in 15 mL
185 tubes and incubated in the dark at room temperature for 45 min to carbamidomethylate cysteines.
186 After alkylation, samples were transferred to an Amicon filter (10,000 MWCO, MilliporeSigma),
187 washed twice with 5 mL 0.1 M Tris, pH 7.5, containing 8 M urea by centrifugation at 4,000 × g,

188 then washed twice with 5 mL 0.1 M Tris, pH 7.5, containing 4 M urea, and finally twice with 5 mL
189 100 mM ammonium bicarbonate (Honeywell). After the last wash step, retentate was reduced to
190 less than 1 mL by spin concentration.

191 *In-Solution Trypsin Proteolytic Cleavage* — MS Grade Trypsin Protease (Pierce) was dissolved
192 to obtain a 20 µg/mL stock in 100 mM ammonium bicarbonate. Trypsin digestion of processed *M.*
193 *smegmatis* lysates was performed in an Amicon filter unit at an approximately 1500:1
194 protein:trypsin ratio overnight at 37 $^{o}$C. Upon digestion, peptides were recovered by centrifugation
195 at 4,000 × g for 10 minutes, followed by the addition of 0.5 M NaCl. Eluted samples were dried
196 under vacuum in a SpeedVac centrifuge (Thermo Fisher).

197 *$TiO_2$ Enrichment of Phosphopeptides* — Phosphopeptides were enriched from tryptic peptide
198 samples using Titansphere $TiO_2$ beads (5 µm; GL Sciences) in buffer conditions that limited acid
199 hydrolysis of phosphoarginine (Thingholm et al., 2006; Schmidt et al., 2014; Trentini et al., 2016).
200 5 mg $TiO_2$ resin was resuspended in 1 mL Binding Buffer (300 mg/mL lactic acid, 12.5% acetic
201 acid, 60% acetonitrile, 0.2% heptafluorobutyric acid, pH 4 with $NH_4OH$). Lyophilized peptide
202 samples were re-dissolved in the $TiO_2$ suspension, incubated for 35 min at 20 $^{o}$C with gentle
203 agitation, then transferred to graphite spin columns (Thermo Fisher). Unbound peptides were
204 removed by a first wash step with 150 µL Binding Buffer and spun at 2000 × g for 1 min, followed
205 by three wash steps using 400 µL of Wash Solution A (200 mg/ml lactic acid, 75% acetonitrile,
206 2% trifluoroacetic acid, 2% heptafluorobutyric acid), Wash Solution B (200 mg/mL lactic acid, 75%
207 acetonitrile, 10% acetic acid, 0.1% heptafluorobutyric acid, pH 4 with $NH_4OH$), Wash Solution C
208 (80% acetonitrile, 10% acetic acid), respectively. The resin was then incubated with 100 µL
209 Elution Solution 1 (1% $NH_4OH$, 30 mM ammonium phosphate) and Elution Solution 2 (1.25%
210 $NH_4OH$ in 50% acetonitrile) for 15 minutes each. The eluate containing phosphopeptides was
211 collected by centrifugation after each incubation. To remove salts, samples were desalted using

212 a HyperSep C18 column (Thermo Scientific). Samples were lyophilized and stored at -80 $^{o}$C prior
213 to mass spectrometry.

214 *Liquid Chromatography - Tandem Mass Spectrometry (LC-MS/MS)* — Lyophilized and desalted
215 tryptic digests were resuspended in 20 µL 0.5% acetic acid (pH 4.5). An Orbitrap Eclipse MS
216 (Thermo Scientific) coupled with an Ultimate 3000 nano-LC system and a FAIMS Pro Interface
217 (Thermo Scientific) was used for the LC-MS/MS analysis. Peptide samples were first loaded onto
218 a trap column (PepMap C18; 2 cm x 100 µm I.D.) and afterwards separated at a flow rate of 300
219 nL/min on an analytical column (PepMap C18, 3.0 µm; 10 cm x 75 mm I.D.; Thermo Scientific). A
220 binary buffer system (buffer A, 0.1% formic acid in water; buffer B, 0.1% formic acid in acetonitrile)
221 with a 165-min gradient (1% to 25% buffer B over 125 min; 25% to 32% buffer B in 10 min, then
222 to 95% buffer B over 3 min; back to 1% B in 5 min, and stay equilibration at 1% B for 20 min) was
223 utilized. To achieve field asymmetric ion mobility spectrometry (FAIMS) separation, multiple CVs
224 (-45, -60 and -80) were applied. For all experiments, the survey scans (MS1) were acquired over
225 a mass range of 375-1500 m/z at a resolution of 60,000 in the Orbitrap. Isolation of precursors
226 was done with a width of 1.6 m/z for MS/MS acquisition. They were subsequently fragmented with
227 higher energy collisional dissociation (HCD) using 30% collision energy with a maximum injection
228 time of 100 ms, and collected in Orbitrap at 15,000 resolution. The dynamic exclusion was set to
229 60 s, and was shared across different FAIMS experiments. LS-MS/MS data was collected in
230 independent biological triplicates.

231 *Mass Spectrometry Data Analysis* — Proteomic analysis was performed in the Proteome
232 Discoverer software suite (version 2.2, Thermo Fisher). Raw data was searched against the *M.*
233 *smegmatis* (strain ATCC 700084 / MC$^2$ 155) UniProt Reference Proteome (Proteome ID
234 UP000000757; 6,602 entries in total) using Sequest HT (University of Washington and Thermo
235 Fisher) (Eng et al., 1994). Iodoacetamide-mediated cysteine carbamidomethylation was set as a
236 static modification, while methionine oxidation and phosphorylation of arginine, serine, threonine,

237 tyrosine and histidine residues were entered as dynamic modifications. Complete trypsinization
238 with a maximum of two missed cleavages was allowed. Precursor mass tolerance was set at 10
239 ppm, while allowing fragment ions to have a mass deviation of 0.02 Da for the HCD data.
240 Validation of peptide-spectrum matches (PSM) based on q-value was done using Percolator -
241 with target false discovery rates (FDR) of 1% and 5% for stringent and relaxed validation,
242 respectively. The false discovery rate of high confidence protein and peptide identification was at
243 1%. Localization probability of phosphopeptide hits were analyzed using the PhosphoRS (ptmRS)
244 node of the Proteome Discoverer software (Taus et al., 2011). Only modifications with a
245 PhosphoRS (ptmRS) score ≥75% were selected.

246 *Gene Ontology Annotation Analysis* — Gene Ontology (GO) annotation was performed using the
247 OmicsBox software suite (Conesa et al., 2005). Homologs were identified by BLAST against the
248 SwissProt/UniProt database (Bairoch et al., 1997), with a minimum expectation value $10^{-3}$.
249 Homolog annotations were compiled from the InterPro database (Hunter et al., 2009). BLAST and
250 InterPro results were used to generate GO terms in terms of biological process. For proteins still
251 unannotated, direct Uniprot BLAST was performed, and tentative assignment of functional group
252 was based on those of obtained homologs.

253 *Sequence and Structural Analysis of Arginine Phosphosites* — For alignment analysis, a 21-mer
254 sequence was obtained containing 10 residues flanking each side of phosphoarginine. All 21-
255 mers were then aligned in BioEdit (Hall, 1999). Structural information, if available, was obtained
256 from Protein Data Bank (PDB) (Berman et al, 2000). Predicted structures were retrieved using
257 neural network-based machine learning model, AlphaFold (version 2, Jumper et al., 2021) and
258 molecular images were prepared in PyMOL (version 2.5.2, Schrödinger). Protein BLAST (Basic
259 Local Alignment Search Tool) of arginine-phosphorylated proteins was performed in the NCBI
260 suite (Gish and States, 1994). Analysis was restricted to the Corynebacterineae suborder.

261 Sequences were aligned using Clustal Omega (Sievers et al., 2011), and conservation and
262 consensus values obtained in Jalview (Waterhouse et al., 2009).

263 *Data Availability* — The mass spectrometry data from this publication have been submitted to the
264 ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE
265 partner repository (Perez-Riverol et al., 2019), and assigned the identifier PXD032083.

266

267 **Results**

268 ***Phosphoarginine Binding Sites in the ClpC1 NTD Are Conserved Across Actinobacteria –***
269 ClpC1 is an essential enzyme in *Mtb* and *Msm* (Sassetti et al., 2003; DeJesus et al., 2017), yet
270 few specific cellular roles or proteolytic substrates of the ClpC1P1P2 protease are known. To gain
271 insight into its cellular function, we began by examining ClpC1 sequence conservation across the
272 phylum Actinobacteria by constructing an alignment of 1,195 actinobacterial orthologs (**Fig. 1A**).
273 Like most other type-II Clp unfoldases, ClpC1 possesses a ~150 amino acid N-terminal domain
274 (NTD), which likely participates in regulation and substrate recognition, but lacks mechanical
275 function or a direct role in catalysis (Schlothauer et al, 2003; Marsee et al., 2018). Surprisingly,
276 the NTD is as well or better conserved than the D1 and D2 AAA+ ATPase rings, suggesting that
277 the NTD has important and conserved functions across Actinobacteria. This may help explain
278 why naturally occurring antibiotics that target and dysregulate ClpC1 have evolved to specifically
279 bind the NTD (Kirstein et al., 2009; Schmitt et al., 2011; Vasudevan et al., 2013; Gavrish et al.,
280 2014; Bürstner et al., 2015; Gao et al., 2015; Jung et al., 2017).

281 In *Bacillus subtilis* and other Firmicutes, ClpCP recognizes phosphorylated arginine
282 residues as degradation signals via twin binding sites on opposite ends of the ClpC NTD (Trentini
283 et al., 2016). Examination of actinobacterial ClpC1 sequences revealed that the residues known
284 to be important for phosphoarginine binding in *B. subtilis* are conserved across actinobacterial

285 orthologs (**Fig. 1B**). We found that we could readily model bound phosphoarginines in an existing
286 structure of the *M. tuberculosis* ClpC1 NTD (Wolf et al., 2020), with minor side chains
287 rearrangements, based on the binding mode observed to the *B. subtilis* ClpC NTD (Trentini et al.,
288 2016) (**Fig. 1C,D**). Finally, we directly assessed binding of pArg to purified *Msm* ClpC1$^{NTD}$ by
289 microscale thermophoresis (**Fig. 1E**). Phosphoarginine bound with an affinity of 1.4 μM, similar to
290 the ~5 μM affinity previously reported to the *Mtb* ClpC1 NTD (Weinhäupl et al., 2018). The ability
291 of the ClpC1$^{NTD}$ to bind pArg and the conservation of pArg binding modules across Actinobacteria
292 provides strong indirect evidence for the existence of this post-translational modification in this
293 phylum.

294 ***Identification of Phosphoarginine Modifications in Mycolicibacterium smegmatis –*** To test
295 for the physiological existence of arginine phosphorylation, we probed dot blots of several
296 bacterial lysates with a phosphoarginine-specific antibody (**Fig. 1F**) (Fuhrmann et al., 2015). In
297 line with prior studies, signal was detected in *B. subtilis* lysate (Fuhrmann et al., 2015), but not in
298 lysate from *E. coli*, which lacks arginine phosphorylation (Potel et al., 2018, Fu et al., 2020). We
299 also observed a positive immunoblot reaction in lysates from *Msm* and a second actinobacterium,
300 *Corynebacterium jeikeium*. These observations suggest that pArg modifications occur at least
301 within the suborder Corynebacterineae, which encompasses *Corynebacterium*,
302 *Mycolicibacterium*, and *Mycobacterium*.

303 To determine which specific cellular proteins carry pArg modifications in *Msm*, we
304 employed an unbiased shotgun proteomics approach. Phosphoarginine is acid-labile and has a
305 short half-life in the acidic TFA-containing solvent systems typically used for proteomic sample
306 preparation and LS-MS/MS (Schmidt et al., 2014). Optimized protocols for arginine
307 phosphoproteomics have been reported that utilize alternative solvent systems at pH ≥ 4 for most
308 steps (Schmidt et al., 2014; Trentini et al., 2016; Trentini et al., 2018). We adapted these methods
309 to our workflow for sample preparation, phosphopeptide enrichment, and LC-MS/MS, thereby

310 minimizing phosphoarginine hydrolysis (**Fig. 2A**). Since stress conditions such as heat shock
311 upregulated the occurrence of this modification in *B. subtilis* (Schmidt et al., 2014), we analyzed
312 lysates from *Msm* cultures grown at either normal growth temperature (37 $^{o}$C) or 50 $^{o}$C for 21
313 hours.

314 A similar number (~3000) of proteins were observed in samples from normal growth (3147,
315 3004, and 3145 proteins in the three replicates, respectively) and heat stress conditions (2969,
316 2724 and 2899 proteins) (**Fig. 2B**). This suggests that at this level of stress, compensatory stress
317 responses do not involve a complete shutdown of translational machinery or dramatic
318 upregulation of proteolytic activity. Typically, a major part of stress response to heat shock is the
319 increased expression of heat shock proteins or chaperones. We evaluated the differential levels
320 of observed proteins based on Sequest HT scores. As shown in **Fig. 2C,** several proteins were
321 observed at significantly higher levels ($p$-value ≤ 0.05) in heat shock than in normal growth
322 conditions. These included Spore protein (Msmeg_5611), which belongs to the Hsp20 small heat
323 shock protein family and was enriched over 2-fold upon heat shock. Other chaperone proteins
324 enriched in heat stressed samples included DnaJ1, DnaK, and the SecB-like chaperone SmegB.
325 We also performed comparative GO annotation analysis of strongly enriched/depleted proteins,
326 whose levels differed significantly and by a magnitude of at least two-fold between conditions
327 (**Supplemental Figure S1; Supplemental Table S1**). Proteins involved in amino acid, lipid and
328 noncanonical metabolism, along with redox proteins, were strongly depleted during heat stress.
329 Spore protein was the only stress response protein strongly enriched under heat shock; no protein
330 involved in stress response was strongly depleted. Taken together, this differential expression
331 suggests the activation of canonical heat shock response in these samples.

332 Using Proteome Discoverer we identified arginine-phosphorylated sites in proteins from
333 these samples. Only peptides with a phosphosite localization probability (PhosphoRS/ptmRS
334 score) ≥75% were selected (Bäsell et al., 2014). We unambiguously localized six phosphoarginine
335 sites in six different proteins (**Fig. 3A; Supplemental Table S2; Supplemental Figure S2**). Most

336 pArg sites were observed in multiple replicates, in both stressed and unstressed samples (**Fig.**
337 **3A**). The existence of phosphoarginine modifications in our samples implies the existence of an
338 unidentified *Msm* pArg kinase.

339 ***Functional Characterization of Arginine-Phosphorylated Proteins –*** We examined the known
340 or predicted functions of the six proteins observed to carry pArg modifications and found that they
341 have diverse functions with no obvious common physiological role (**Fig. 3A**). PhoU1 is a central
342 regulator of the SenX3-RegX3 two-component system responsible for uptake of inorganic
343 phosphate ($P_i$) during phosphate starvation (Brokaw et al., 2017). This raises the possibility that
344 pArg modifications are linked to $P_i$ availability. Two other pArg-bearing targets were metabolic
345 enzymes: the multifunctional 2-oxoglutarate metabolism enzyme Kgd, involved in the tricarboxylic
346 acid cycle (Wagner et al., 2011), and the glycerol dehydratase large subunit (Msmeg_1547),
347 which potentially contributes to the catabolic pathway involving the glycerol dehydration reaction
348 which yields 3-hydroxypropanal in the presence of adenosylcobalamin coenzyme (**Fig. 3A;**
349 **Supplemental Figure S2**) (Liu et al., 2010; Nasir et al., 2020). pArg may thus play a role in
350 regulating cellular metabolic pathways. Another phosphosite was detected in the transmembrane
351 xanthine/uracil permease (Msmeg_1293), a nucleobase transporter (**Figs. 3A and 3C**). A site
352 was found on the hemerythrin domain-containing protein Msmeg_0907, which belongs to a class
353 of $O_2$-binding proteins involved in signal transduction, response to $H_2O_2$, oxygen sensing and nitric
354 oxide reduction (Isaza et al., 2006; Li et al. 2015; Lo et al., 2016; Ma et al., 2020). Finally, a site
355 was observed on Msmeg_0879, a small (48 aa) uncharacterized protein predicted to be
356 predominantly disordered (**Figs. 3A and B**). Homologs of Msmeg_0879 were detected in close
357 relatives of *Msm*, but are not widely conserved in actinobacteria.

358 ***Structural and Physicochemical Analysis of Arginine Phosphorylation Sites –*** We assessed
359 whether arginine phosphorylation occurred at sites with particular physicochemical properties.
360 We aligned 21-residue sequence segments centered on each unique phosphoarginine site (**Fig.**

361 **4A**) but observe no clear consensus motif. Ala, Leu, Val and Gly appear to be common in flanking
362 positions, although we note that these are the four most abundant amino acids in the *M.*
363 *smegmatis* proteome (13%, 10%, 9%, 9% of total, respectively) (**Supplemental Table S3**). The
364 sequence diversity surrounding pArg positions suggests that target discrimination is guided by
365 characteristics other than primary sequence.

366 We next examined structural characteristics of phosphorylated positions, based on
367 predicted structures generated by AlphaFold2 (Jumper et al., 2021), except for the small protein
368 Msmeg_0879 for which structure prediction failed. (Notably, the AlphaFold2 prediction for kgd
369 was virtually identical to its reported X-ray structure (PDB ID: 2XT6) with RMSD ~0.5 A in the
370 region of the phosphosite. (Wagner et al., 2011)) As expected for this charged residue, arginine
371 phosphosites were solvent exposed (**Supplemental Figure S3**), but were located on structured
372 elements rather than loops. Interestingly, in all structures, arginine phosphosites occurred
373 proximal to the beginning or the end of an alpha helix (**Figs. 4B-F**), which may reflect a structural
374 constraint important for recognition by an arginine kinase.

375 Finally, we assessed the sequence conservation of arginine at these positions. We aligned
376 the arginine-phosphorylated *Msm* proteins with homologs within the Corynebacterineae suborder,
377 and analyzed the positional conservation of arginine at the respective positions. As shown in **Fig.**
378 **5**, conservation varies. The phosphosite arginine was well conserved in Kgd (99.8%),
379 Msmeg_0907 (79%), and Msmeg_0879 (62%) (**Figs 5A-C**). Conservation was lower in
380 Msmeg_1547 (34.1%) and PhoU1 (32.63%) (**Figs. 5D and 5E**). Arg was rarely present in this
381 position in homologs of Msmeg_1293 (0.23%) (**Fig. 5F**). In cases where the phosphorylated
382 arginine was well conserved, it may indicate that phosphorylation at this position is a conserved
383 phenomenon with functional or regulatory significance.

384

385 **Discussion**

386 Protein phosphorylation is a ubiquitous mechanism of signal propagation and pathway
387 regulation in bacteria (Macek et al., 2019). While examples of His, Asp, Ser, Thr, and Tyr
388 phosphorylation are widespread (Chow et al., 1994; Canova et al., 2008; Prisic et al., 2010;
389 Nakedi et al., 2015), the importance of arginine phosphorylation has become apparent only
390 recently. A growing body of data links pArg modifications to protein quality control pathways and
391 spore germination in Firmicutes, including *B. subtilis* (Fuhrmann et al., 2009; Elsholz et al., 2012;
392 Schmidt et al., 2014; Fuhrmann et al., 2016; Trentini et al., 2016; Junker et al., 2018; Weinhäupl
393 et al., 2018; Zhou et al., 2019). However, it has remained unclear how prevalent pArg
394 modifications are in other bacterial phyla. Here we provide direct evidence for the existence of
395 arginine phosphorylation in *Mycolicibacterium smegmatis*, an actinobacterium.

396 In *B. subtilis* and *S. aureus*, this modification is upregulated by stress (Fuhrmann et al.,
397 2009; Elsholz et al., 2012; Schmidt et al., 2014; Fuhrmann et al., 2016; Trentini et al., 2016; Junker
398 et al., 2018; Weinhäupl et al., 2018; Zhou et al., 2019). Hence, we performed our study by
399 examining pArg levels in cells grown under heat stress and normal conditions. While we saw
400 expected increases in levels of some heat shock proteins overall, arginine phosphorylation
401 abundance was not altered by heat shock, as occurs in firmicutes. It is possible that mycobacterial
402 pArg levels are linked to different stress stimuli that were not tested here. Alternatively, pArg may
403 have no relationship to the mycobacterial stress response, and may instead play a targeted
404 regulatory role.

405 The six arginine phosphosites localized here were far fewer than those identified in other
406 bacterial studies. Two independent studies in *B. subtilis* observed 121 and 217 sites in 87 and
407 134 proteins, respectively (Elsholz et al., 2012; Schmidt et al., 2014), while in *S. aureus* 207 sites
408 were identified in 126 proteins (Junker et al., 2018). On its face, the small number of pArg
409 modifications found in *Msm* argues against a proteome-wide quality control role, and appears
410 more consistent with targeted regulation of select proteins and processes. However, we note that
411 phosphoproteomic studies in Firmicutes utilized deletion strains lacking known arginine

412 phosphatases ($^{Bs}$YwlE and $^{Sa}$PtpB), which elevate levels of pArg-bearing proteins. No arginine
413 phosphatase has yet been identified in *Msm* or *Mtb*, thus our approach utilized wild-type *Msm*. If
414 a mycobacterial arginine phosphatase is eventually identified, it would be interesting to observe
415 the effect of its knockout or knockdown on pArg levels.

416 For arginine phosphorylation to be useful to the cell, it must presumably be applied
417 selectively through the regulated activity of arginine kinases. Many protein kinases recognize
418 substrates through characteristic sequence motifs surrounding the phosphosite. By contrast, our
419 examination of mycobacterial arginine phosphosites revealed no clearly enriched consensus
420 sequence around pArg. Prior analysis of Arg phosphosites in *B. subtilis* similarly revealed no
421 evident consensus sequence (Schmidt et al., 2014). This degeneracy suggests either that distal
422 interactions guide *Msm* arginine kinases to phosphosites or that other characteristics of target
423 proteins guide substrate selection. Supporting this latter possibility, we noted that all
424 phosphorylated arginine residues (for which structural models could be obtained) were found near
425 one end of an alpha helix, which may indicate a structural feature recognized by an arginine
426 kinase.

427 What role does arginine phosphorylation play in mycobacterial cells? In *B. subtilis*, pArg
428 tagging helps enforce protein quality control by marking misfolded proteins for destruction by
429 ClpCP (Elsholz et al., 2012; Schmidt et al., 2014; Trentini et al., 2016). Prior studies have
430 demonstrated that mycobacterial ClpC1P1P2 can degrade model substrates *in vitro* bearing pArg
431 modifications (Weinhäupl et al., 2018). The pArg-bearing proteins identified here may therefore
432 be recognized as proteolytic substrates by ClpC1P1P2. Alternatively, pArg may modulate some
433 aspect of the targets' function or interactions with other proteins. We observed pArg on PhoU1,
434 which helps regulate $P_i$ uptake by inhibiting the activity of the SenX3-RegX3 two-component
435 system when $P_i$ is readily available (Brokaw et al., 2017). Additionally, the signal transduction
436 protein Msmeg_0907, metabolic proteins (Kgd and Msmeg_1547), a transmembrane
437 xanthine/uracil transport protein (Msmeg_1293) and an uncharacterized protein Msmeg_0879

438 were found to be phosphorylated. In terms of essentiality, a prior study showed that the *Mtb*
439 homolog of Kgd is essential for viability (Sassetti et al., 2003); while another report showed that
440 PhoU1 is jointly essential with the PhoU2 protein (Msmeg_1605) for *in vitro Msm* growth (Brokaw
441 et al., 2017). More work is required to determine whether pArg modifications regulate these
442 pathways, and whether this modification plays an essential role in mycobacterial physiology.
443 This study lays the groundwork for future efforts to expound on the roles of
444 phosphoarginine modifications in mycobacteria. The immediate impediment to further
445 understanding this system is the absence of identifiable orthologs of known arginine kinases or
446 phosphatases. Nevertheless, the lack of consensus motif in the identified phosphosites predicts
447 a promiscuous kinase that targets cellular proteins, in a manner independent of a specific primary
448 sequence. Future work will be required to decipher how arginine phosphorylation is regulated,
449 and how it contributes to the physiology of mycobacteria and other actinobacterial species.
450

451 **SUPPLEMENTAL DATA**

452 This article contains supplemental data.
453

454 **ACKNOWLEDGEMENTS**

455 We thank members of the Schmitz lab for advice; VP, AJL, PCB, MP, AD, PAO and OU for
456 comments on the manuscript; and the Parashar lab for use of instrumentation. For technical help
457 with mass spectrometry, we thank Dr Yanbao Yu and the University of Delaware Mass
458 Spectrometry Facility. Sortase A was a kind gift from S Novo and JM Fox. The anti-pArg antibody
459 was a kind gift from Dr. Paul Thompson, UMass Medical School.
460

461 **Funding and Additional Information**

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

466

467 **Author Contributions**

468 E. O. and K. S. designed research; E. O. and H. A. performed research; E. O. and K. S. analyzed
469 data; E. O. and K. S. wrote the paper.

470

471 **References**

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

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

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

482 Baker, T. A., and Sauer, R. T. (2012) ClpXP, an ATP-powered unfolding and protein-degradation
483 machine. *Biochimica et biophysica acta* **1823(1):** 15–28.

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

487 Bäsell, K., Otto, A., Junker, S., Zühlke, D., Rappen, G. M., Schmidt, S., Hentschker, C., Macek,
488 B., Ohlsen, K., Hecker, M., and Becher, D. (2014) The phosphoproteome and its
489 physiological dynamics in Staphylococcus aureus. *International journal of medical*
490 *microbiology: IJMM* **304(2):** 121–132.
491 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N.,
492 & Bourne, P. E. (2000) The Protein Data Bank. *Nucleic acids research* **28(1):** 235–242.
493 Brokaw, A. M., Eide, B. J., Muradian, M., Boster, J. M., & Tischler, A. D. (2017) *Mycobacterium*
494 *smegmatis* PhoU Proteins Have Overlapping Functions in Phosphate Signaling and Are
495 Essential. *Frontiers in microbiology* **8:** 2523.
496 Brötz-Oesterhelt, H., Beyer, D., Kroll, H. P., Endermann, R., Ladel, C., Schroeder, W., Hinzen,
497 B., Raddatz, S., Paulsen, H., Henninger, K., Bandow, J. E., Sahl, H. G., and Labischinski,
498 H. (2005) Dysregulation of bacterial proteolytic machinery by a new class of
499 antibiotics. *Nature medicine* **11(10):** 1082–1087.
500 Bürstner, N., Roggo, S., Ostermann, N., Blank, J., Delmas, C., Freuler, F., Gerhartz, B., Hinniger,
501 A., Hoepfner, D., Liechty, B., Mihalic, M., Murphy, J., Pistorius, D., Rottmann, M., Thomas,
502 J. R., Schirle, M., and Schmitt, E. K. (2015) Gift from Nature: Cyclomarin A Kills
503 Mycobacteria and Malaria Parasites by Distinct Modes of Action. *Chembiochem : a*
504 *European journal of chemical biology* **16(17):** 2433–2436.
505 Canova, M. J., Veyron-Churlet, R., Zanella-Cleon, I., Cohen-Gonsaud, M., Cozzone, A. J., Becchi,
506 M., Kremer, L., and Molle, V. (2008) The *Mycobacterium tuberculosis* serine/threonine
507 kinase PknL phosphorylates Rv2175c: mass spectrometric profiling of the activation loop
508 phosphorylation sites and their role in the recruitment of Rv2175c. *Proteomics* **8(3):** 521–
509 533.
510 Chow, K., Ng, D., Stokes, R., and Johnson, P. (1994) Protein tyrosine phosphorylation in
511 *Mycobacterium tuberculosis*. *FEMS microbiology letters* **124(2):** 203–207.

512 Compton, C. L., Schmitz, K. R., Sauer, R. T., and Sello, J. K. (2013) Antibacterial activity of and
513 resistance to small molecule inhibitors of the ClpP peptidase. *ACS chemical biology* **8(12):**
514 2669–2677.
515 Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005) Blast2GO:
516 a universal tool for annotation, visualization and analysis in functional genomics
517 research. *Bioinformatics (Oxford, England)* **21(18):** 3674–3676.
518 Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo
519 generator. *Genome research* **14(6):** 1188–1190.
520 DeJesus, M.A., Gerrick, E.R., Xu, W., Park, S.W., Long, J.E. Boutte, C.C., Rubin, E.J.,
521 Schnappinger, D., Ehrt, S., Fortune, S.M., Sassetti, C.M., and Ioerger, T.R. (2017)
522 Comprehensive Essentiality Analysis of the *Mycobacterium tuberculosis* Genome via
523 Saturating Transposon Mutagenesis. *mBio* **8(1):** e02133-16.
524 Eddy S. R. (1995) Multiple alignment using hidden Markov models. *Proceedings. International*
525 *Conference on Intelligent Systems for Molecular Biology* **3:** 114–120.
526 Elsholz, A. K., Turgay, K., Michalik, S., Hessling, B., Gronau, K., Oertel, D., Mäder, U., Bernhardt,
527 J., Becher, D., Hecker, M., and Gerth, U. (2012) Global impact of protein arginine
528 phosphorylation on the physiology of *Bacillus subtilis*. *Proceedings of the National*
529 *Academy of Sciences of the United States of America* **109(19):** 7451–7456.
530 Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass
531 spectral data of peptides with amino acid sequences in a protein database. *Journal of the*
532 *American Society for Mass Spectrometry* **5(11):** 976–989.
533 Famulla, K., Sass, P., Malik, I., Akopian, T., Kandror, O., Alber, M., Hinzen, B., Ruebsamen-
534 Schaeff, H., Kalscheuer, R., Goldberg, A. L., and Brötz-Oesterhelt, H. (2016)
535 Acyldepsipeptide antibiotics kill mycobacteria by preventing the physiological functions of
536 the ClpP1P2 protease. *Molecular microbiology* **101(2)**: 194–209.

537 Fu, S., Fu, C., Zhou, Q., Lin, R., Han, O., Wang, M., Sun, Y., Liu, Y., and Zhao, Y.
538 (2020) Widespread arginine phosphorylation in human cells—a novel protein PTM
539 revealed by mass spectrometry. *Sci. China Chem.* **63:** 341–346.
540 Fuhrmann, J., Schmidt, A., Spiess, S., Lehner, A., Turgay, K., Mechtler, K., Charpentier, E., and
541 Clausen, T. (2009) McsB is a protein arginine kinase that phosphorylates and inhibits the
542 heat-shock regulator CtsR. *Science (New York, N.Y.)* **324(5932):** 1323–1327.
543 Fuhrmann, J., Subramanian, V., and Thompson, P. R. (2015) Synthesis and Use of a
544 Phosphonate Amidine to Generate an Anti-Phosphoarginine-Specific
545 Antibody. *Angewandte Chemie (International ed. in English)* **54(49):** 14715–14718.
546 Fuhrmann, J., Subramanian, V., Kojetin, D. J., and Thompson, P. R. (2016) Activity-Based
547 Profiling Reveals a Regulatory Link between Oxidative Stress and Protein Arginine
548 Phosphorylation. *Cell chemical biology* **23(8):** 967–977.
549 Gao, W., Kim, J. Y., Anderson, J. R., Akopian, T., Hong, S., Jin, Y. Y., Kandror, O., Kim, J. W.,
550 Lee, I. A., Lee, S. Y., McAlpine, J. B., Mulugeta, S., Sunoqrot, S., Wang, Y., Yang, S. H.,
551 Yoon, T. M., Goldberg, A. L., Pauli, G. F., Suh, J. W., Franzblau, S. G., and Cho, S. (2015)
552 The cyclic peptide ecumicin targeting ClpC1 is active against *Mycobacterium tuberculosis*
553 in vivo. *Antimicrobial agents and chemotherapy* **59(2):** 880–889.
554 Gavrish, E., Sit, C. S., Cao, S., Kandror, O., Spoering, A., Peoples, A., Ling, L., Fetterman, A.,
555 Hughes, D., Bissell, A., Torrey, H., Akopian, T., Mueller, A., Epstein, S., Goldberg, A.,
556 Clardy, J., and Lewis, K. (2014) Lassomycin, a ribosomally synthesized cyclic peptide,
557 kills *Mycobacterium tuberculosis* by targeting the ATP-dependent protease
558 ClpC1P1P2. *Chemistry & biology* **21(4):** 509–518.
559 Gerth, U., Kock, H., Kusters, I., Michalik, S., Switzer, R. L., & Hecker, M. (2008) Clp-dependent
560 proteolysis down-regulates central metabolic pathways in glucose-starved Bacillus
561 subtilis. *Journal of bacteriology* **190(1):** 321–331.

562 Glasgow, J. E., Salit, M. L., and Cochran, J. R. (2016) *In vivo* site-specific protein tagging with
563 diverse amines using an engineered sortase variant. *Journal of the American Chemical*
564 *Society* **138 (24):** 7496-7499
565 Gopal, P., Sarathy, J. P., Yee, M., Ragunathan, P., Shin, J., Bhushan, S., Zhu, J., Akopian, T.,
566 Kandror, O., Lim, T. K., Gengenbacher, M., Lin, Q., Rubin, E. J., Grüber, G., and Dick, T.
567 (2020) Pyrazinamide triggers degradation of its target aspartate decarboxylase. *Nature*
568 *communications* **11(1):** 1661.
569 Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger ,T.R., Akerley, B.J., and Sassetti, C.M.
570 (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial
571 growth and cholesterol catabolism. *PLoS Pathog* **7(9):** e1002251.
572 Hall, T.A. (1999) BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis
573 Program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95-98.
574 Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U.,
575 Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly,
576 E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., and Yeats, C. (2009) InterPro: the
577 integrative protein signature database. *Nucleic acids research* **37(Database issue):**
578 D211–D215.
579 Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., &
580 Dunker, A. K. (2004) The importance of intrinsic disorder for protein
581 phosphorylation. *Nucleic acids research* **32(3):** 1037–1049.
582 Isaza, C. E., Silaghi-Dumitrescu, R., Iyer, R. B., Kurtz, D. M., Jr, & Chan, M. K. (2006) Structural
583 basis for $O_2$ sensing by the hemerythrin-like domain of a bacterial chemotaxis protein:
584 substrate tunnel and fluxional N terminus. *Biochemistry* **45(30):** 9023–9031.
585 Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P., & Breitsprecher,
586 D. (2014) MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of*
587 *Molecular Structure* **1077:** 101-113.

588 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O.,Tunyasuvunakool,
589 K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S., Ballard, A. J.,
590 Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Hassabis, D.
591 (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596(7873):**
592 583–589.
593 Jung, I. P., Ha, N. R., Kim, A. R., Kim, S. H., and Yoon, M. Y. (2017) Mutation analysis of the
594 interactions between *Mycobacterium tuberculosis* caseinolytic protease C1 (ClpC1) and
595 ecumicin. *International journal of biological macromolecules* **101:** 348–357.
596 Junker, S., Maaβ, S., Otto, A., Michalik, S., Morgenroth, F., Gerth, U., Hecker, M., and Becher,
597 D. (2018) Spectral Library Based Analysis of Arginine Phosphorylations
598 in *Staphylococcus aureus*. *Molecular & cellular proteomics: MCP* **17(2):** 335–348.
599 Kirstein, J., Hoffmann, A., Lilie, H., Schmidt, R., Rübsamen-Waigmann, H., Brötz-Oesterhelt, H.,
600 Mogk, A., and Turgay, K. (2009) The antibiotic ADEP reprogrammes ClpP, switching it
601 from a regulated to an uncontrolled protease. *EMBO molecular medicine* **1(1):** 37–49.
602 Lavey, N. P., Coker, J. A., Ruben, E. A., and Duerfeldt, A. S. (2016) Sclerotiamide: The First Non-
603 Peptide-Based Natural Product Activator of Bacterial Caseinolytic Protease P. *Journal of*
604 *natural products* **79(4):** 1193–1197.
605 Leodolter, J., Warweg, J., and Weber-Ban, E. (2015) The *Mycobacterium tuberculosis* ClpP1P2
606 Protease Interacts Asymmetrically with Its ATPase Partners ClpX and ClpC1. *PloS*
607 *one* **10(5):** e0125345.
608 Li, L., MacIntyre, L. W., Ali, T., Russo, R., Koirala, B., Hernandez, Y., & Brady, S. F. (2021)
609 Biosynthetic Interrogation of Soil Metagenomes Reveals Metamarin, an Uncommon
610 Cyclomarin Congener with Activity against *Mycobacterium tuberculosis*. *Journal of natural*
611 *products* **84(4):** 1056–1066.
612 Li, M., Kandror, O., Akopian, T., Dharkar, P., Wlodawer, A., Maurizi, M. R., and Goldberg, A. L.
613 (2016) Structure and Functional Properties of the Active Form of the Proteolytic Complex,

614 ClpP1P2, from *Mycobacterium tuberculosis*. *The Journal of biological chemistry* **291(14):**
615 7465–7476.
616 Li, X., Li, J., Hu, X. *et al.* (2015) Differential roles of the hemerythrin-like proteins
617 of *Mycobacterium smegmatis* in hydrogen peroxide and erythromycin susceptibility. *Sci*
618 *Rep* **5:** 16130
619 Liu, Y., Gallo, A. A., Bajpai, R. K., Chistoserdov, A., Nelson, A., Segura, L., & Xu, W. (2010) The
620 diversity and molecular modelling analysis of $B_{12}$ and $B_{12}$-independent glycerol
621 dehydratases. *International journal of bioinformatics research and applications* **6(5):** 484–
622 507.
623 Lunge, A., Gupta, R., Choudhary, E., and Agarwal, N. (2020) The unfoldase ClpC1
624 of *Mycobacterium tuberculosis* regulates the expression of a distinct subset of proteins
625 having intrinsically disordered termini. *The Journal of biological chemistry* **295(28):** 9455–
626 9473.
627 Ma, Z., Abendroth, J., Buchko, G. W., Rohde, K. H., & Davidson, V. L. (2020) Crystal structure of
628 a hemerythrin-like protein from *Mycobacterium kansasii* and homology model of the
629 orthologous Rv2633c protein of *M. tuberculosis*. *The Biochemical journal* **477(2**): 567–
630 581.
631 Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C., and Mijakovic, I.
632 (2019) Protein post-translational modifications in bacteria. *Nature reviews.*
633 *Microbiology* **17(11):** 651–664.
634 Marsee, J. D., Ridings, A., Yu, T., and Miller, J. M. (2018) *Mycobacterium tuberculosis* ClpC1 N-
635 Terminal Domain Is Dispensable for Adaptor Protein-Dependent Allosteric
636 Regulation. *International journal of molecular sciences* **19(11):** 3651.
637 Moreira, W., Ngan, G. J., Low, J. L., Poulsen, A., Chia, B. C., Ang, M. J., Yap, A., Fulwood, J.,
638 Lakshmanan, U., Lim, J., Khoo, A. Y., Flotow, H., Hill, J., Raju, R. M., Rubin, E. J., and

639 Dick, T. (2015) Target mechanism-based whole-cell screening identifies bortezomib as an
640 inhibitor of caseinolytic protease in mycobacteria. *mBio* **6(3):** e00253-15.
641 Moreno-Cinos, C., Goossens, K., Salado, I. G., Van Der Veken, P., De Winter, H., and Augustyns,
642 K. (2019) ClpP Protease, a Promising Antimicrobial Target. *International journal of*
643 *molecular sciences* **20(9):** 2232.
644 Nagpal, J., Paxman, J. J., Zammit, J. E., Alhuwaider, A., Truscott, K. N., Heras, B., and Dougan,
645 D. A. (2019) Molecular and structural insights into an asymmetric proteolytic
646 complex (ClpP1P2) from *Mycobacterium smegmatis*. *Scientific reports* **9(1):** 18019.
647 Nakedi, K. C., Nel, A. J., Garnett, S., Blackburn, J. M., and Soares, N. C. (2015) Comparative
648 Ser/Thr/Tyr phosphoproteomics between two mycobacterial species: the fast growing
649 *Mycobacterium smegmatis* and the slow growing *Mycobacterium bovis* BCG. *Frontiers in*
650 *microbiology* **6:** 237.
651 Nasir, A., Ashok, S., Shim, J. Y., Park, S., & Yoo, T. H. (2020) Recent Progress in the
652 Understanding and Engineering of Coenzyme $B_{12}$-Dependent Glycerol
653 Dehydratase. *Frontiers in bioengineering and biotechnology* **8**: 500867.
654 Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J.,
655 Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J.,
656 Sachsenberg, T., Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A. F.,
657 and Vizcaíno, J. A. (2019) The PRIDE database and related tools and resources in 2019:
658 improving support for quantification data. *Nucleic acids research* **47(D1):** D442–D450.
659 Potel, C. M., Lin, M. H., Heck, A., and Lemeer, S. (2018) Widespread bacterial protein histidine
660 phosphorylation revealed by mass spectrometry-based proteomics. *Nature*
661 *methods* **15(3):** 187–190.
662 Prisic, S., Dankwa, S., Schwartz, D., Chou, M. F., Locasale, J. W., Kang, C. M., Bemis, G.,
663 Church, G. M., Steen, H., and Husson, R. N. (2010) Extensive phosphorylation with
664 overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein

665 kinases. *Proceedings of the National Academy of Sciences of the United States of*
666 *America* **107(16):** 7521–7526.
667 Raju, R. M., Unnikrishnan, M., Rubin, D. H., Krishnamoorthy, V., Kandror, O., Akopian, T. N.,
668 Goldberg, A. L., and Rubin, E. J. (2012) *Mycobacterium tuberculosis* ClpP1 and ClpP2
669 function together in protein degradation and are required for viability in vitro and during
670 infection. *PLoS pathogens* **8(2):** e1002511.
671 Raju, R.M., Jedrychowski, M.P., Wei, J.R., Pinkham, J.T., Park, A.S., O'Brien, K., Rehren, G.,
672 Schnappinger, D., Gygi S.P., and Rubin, E.J. (2014) Post-translational regulation via Clp
673 protease is critical for survival of *Mycobacterium tuberculosis*. *PLoS Pathog* **10(3):**
674 e1003994.
675 Sassetti, C.M., Boyd, D.H., and Rubin E.J. (2003) Genes required for mycobacterial growth
676 defined by high density mutagenesis. *Molecular microbiology* **48 (1):** 77-84.
677 Sauer, R. T., and Baker, T. A. (2011) AAA+ proteases: ATP-fueled machines of protein
678 destruction. *Annual review of biochemistry* **80:** 587–612.
679 Schlothauer, T., Mogk, A., Dougan, D. A., Bukau, B., and Turgay, K. (2003) MecA, an adaptor
680 protein necessary for ClpC chaperone activity. *Proceedings of the National Academy of*
681 *Sciences of the United States of America* **100(5):** 2306–2311.
682 Schmidt, A., Trentini, D. B., Spiess, S., Fuhrmann, J., Ammerer, G., Mechtler, K., and Clausen,
683 T. (2014) Quantitative phosphoproteomics reveals the role of protein arginine
684 phosphorylation in the bacterial stress response. *Molecular & cellular proteomics:*
685 *MCP* **13(2):** 537–550.
686 Schmitt, E. K., Riwanto, M., Sambandamurthy, V., Roggo, S., Miault, C., Zwingelstein, C., Krastel,
687 P., Noble, C., Beer, D., Rao, S. P., Au, M., Niyomrattanakit, P., Lim, V., Zheng, J., Jeffery,
688 D., Pethe, K., & Camacho, L. R. (2011) The natural product cyclomarin kills
689 *Mycobacterium tuberculosis* by targeting the ClpC1 subunit of the caseinolytic
690 protease. *Angewandte Chemie (International ed. in English)* **50(26)**: 5889–5891.

691 Schmitz, K. R., and Sauer, R. T. (2014) Substrate delivery by the AAA+ ClpX and ClpC1
692 unfoldases activates the mycobacterial ClpP1P2 peptidase. *Molecular*
693 *microbiology* **93(4):** 617–628.
694 Schmitz, K. R., Carney, D. W., Sello, J. K., and Sauer, R. T. (2014) Crystal structure of
695 *Mycobacterium tuberculosis* ClpP1P2 suggests a model for peptidase activation by AAA+
696 partner binding and substrate delivery. *Proceedings of the National Academy of Sciences*
697 *of the United States of America* **111(43):** E4587–E4595.
698 Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H.,
699 Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable
700 generation of high-quality protein multiple sequence alignments using Clustal
701 Omega. *Molecular systems biology* **7:** 539.
702 Singh, G. P. (2015) Association between intrinsic disorder and serine/threonine phosphorylation
703 in *Mycobacterium tuberculosis*. *PeerJ* **3:** e724.
704 States, D. J., & Gish, W. (1994) Combined use of sequence similarity and codon bias for coding
705 region identification. *Journal of computational biology: a journal of computational*
706 *molecular cell biology* **1(1):** 39–50.
707 Suskiewicz, M. J., Hajdusits, B., Beveridge, R., Heuck, A., Vu, L. D., Kurzbauer, R., Hauer, K.,
708 Thoeny, V., Rumpel, K., Mechtler, K., Meinhart, A., and Clausen, T. (2019) Structure of
709 McsB, a protein kinase for regulated arginine phosphorylation. *Nature chemical*
710 *biology* **15(5):** 510–518.
711 Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., and Mechtler, K. (2011)
712 Universal and confident phosphorylation site localization using phosphoRS. *Journal of*
713 *proteome research* **10(12):** 5354–5362.
714 Thingholm, T. E., Jørgensen, T. J., Jensen, O. N., and Larsen, M. R. (2006) Highly selective
715 enrichment of phosphorylated peptides using titanium dioxide. *Nature protocols* **1(4):**
716 1929–1935.

717 Trentini, D. B., Fuhrmann, J., Mechtler, K., & Clausen, T. (2014) Chasing Phosphoarginine
718 Proteins: Development of a Selective Enrichment Method Using a Phosphatase Trap.
719 *Molecular & cellular proteomics: MCP* **13(8):** 1953–1964.
720 Trentini, D. B., Suskiewicz, M. J., Heuck, A., Kurzbauer, R., Deszcz, L., Mechtler, K., and Clausen,
721 T. (2016) Arginine phosphorylation marks proteins for degradation by a Clp
722 protease. *Nature* **539(7627):** 48–53.
723 Vahidi, S., Ripstein, Z. A., Juravsky, J. B., Rennella, E., Goldberg, A. L., Mittermaier, A. K.,
724 Rubinstein, J. L., and Kay, L. E. (2020) An allosteric switch regulates *Mycobacterium*
725 *tuberculosis* ClpP1P2 protease function as established by cryo-EM and methyl-TROSY
726 NMR. *Proceedings of the National Academy of Sciences of the United States of*
727 *America* **117(11):** 5895–5906.
728 Vasudevan, D., Rao, S. P., and Noble, C. G. (2013) Structural basis of mycobacterial inhibition
729 by cyclomarin A. *The Journal of biological chemistry* **288(43):** 30883–30891.
730 Wagner, T., Bellinzoni, M., Wehenkel, A., O'Hare, H. M., & Alzari, P. M. (2011) Functional
731 plasticity and allosteric regulation of α-ketoglutarate decarboxylase in central
732 mycobacterial metabolism. *Chemistry & biology* **18(8):** 1011–1020.
733 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview
734 Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*
735 *(Oxford, England)* **25(9):** 1189–1191.
736 Weinhäupl, K., Brennich, M., Kazmaier, U., Lelievre, J., Ballell, L., Goldberg, A., Schanda, P., and
737 Fraga, H. (2018) The antibiotic cyclomarin blocks arginine-phosphate-induced millisecond
738 dynamics in the N-terminal domain of ClpC1 from *Mycobacterium tuberculosis*. *The*
739 *Journal of biological chemistry* **293(22):** 8379–8393.
740 Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation
741 method for proteome analysis. *Nature methods* **6(5):** 359–362.

742 Wolf, N. M., Lee, H., Choules, M. P., Pauli, G. F., Phansalkar, R., Anderson, J. R., Gao, W., Ren,
743 J., Santarsiero, B. D., Lee, H., Cheng, J., Jin, Y. Y., Ho, N. A., Duc, N. M., Suh, J. W.,
744 Abad-Zapatero, C., and Cho, S. (2019) High-Resolution Structure of ClpC1-Rufomycin
745 and Ligand Binding Studies Provide a Framework to Design and Optimize Anti-
746 Tuberculosis Leads. *ACS infectious diseases* **5(6):** 829–840.
747 Wolf, N. M., Lee, H., Zagal, D., Nam, J. W., Oh, D. C., Lee, H., Suh, J. W., Pauli, G. F., Cho, S.,
748 and Abad-Zapatero, C. (2020) Structure of the N-terminal domain of ClpC1 in complex
749 with the antituberculosis natural product ecumicin reveals unique binding
750 interactions. *Acta crystallographica. Section D, Structural biology* **76(Pt 5):** 458–471.
751 World Health Organization (2021) *Global Tuberculosis Report* (World Health Organization,
752 Geneva)
753 Zhou, B., Semanjski, M., Orlovetskie, N., Bhattacharya, S., Alon, S., Argaman, L., Jarrous, N.,
754 Zhang, Y., Macek, B., Sinai, L., and Ben-Yehuda, S. (2019) Arginine dephosphorylation
755 propels spore germination in bacteria. *Proceedings of the National Academy of Sciences*
756 *of the United States of America* **116(28):** 14228–14237.
757 Ziemski, M., Leodolter, J., Taylor, G., Kerschenmeyer, A., & Weber-Ban, E. (2021) Genome-wide
758 interaction screen for *Mycobacterium tuberculosis* ClpCP protease reveals toxin-antitoxin
759 systems as a major substrate class. *The FEBS journal*, **288(1):** 111–126
760

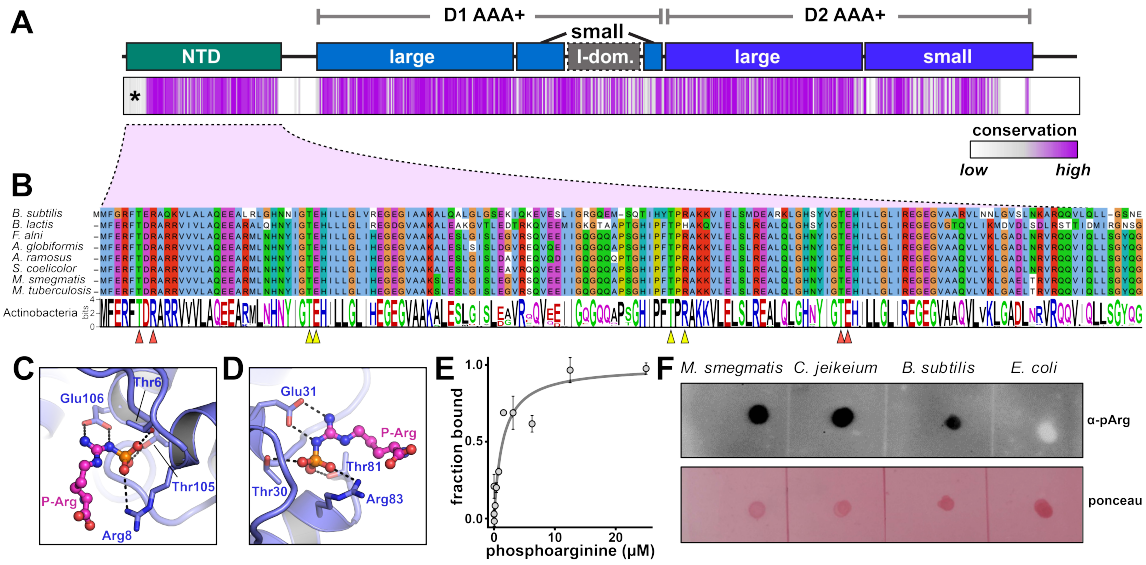



**Figure 1**. **Indirect evidence for phosphoarginine in actinobacteria**. **A**) The domain organization of ClpC1 is shown above a plot of amino acid conservation among actinobacterial ClpC1 orthologs, where conservation at each alignment position is plotted as a purple vertical strip. The apparent 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) (Weinhäupl et al., 2018). 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**) (Wolf et al., 2020). **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 (Fuhrmann et al., 2015) or stained with Ponceau S.

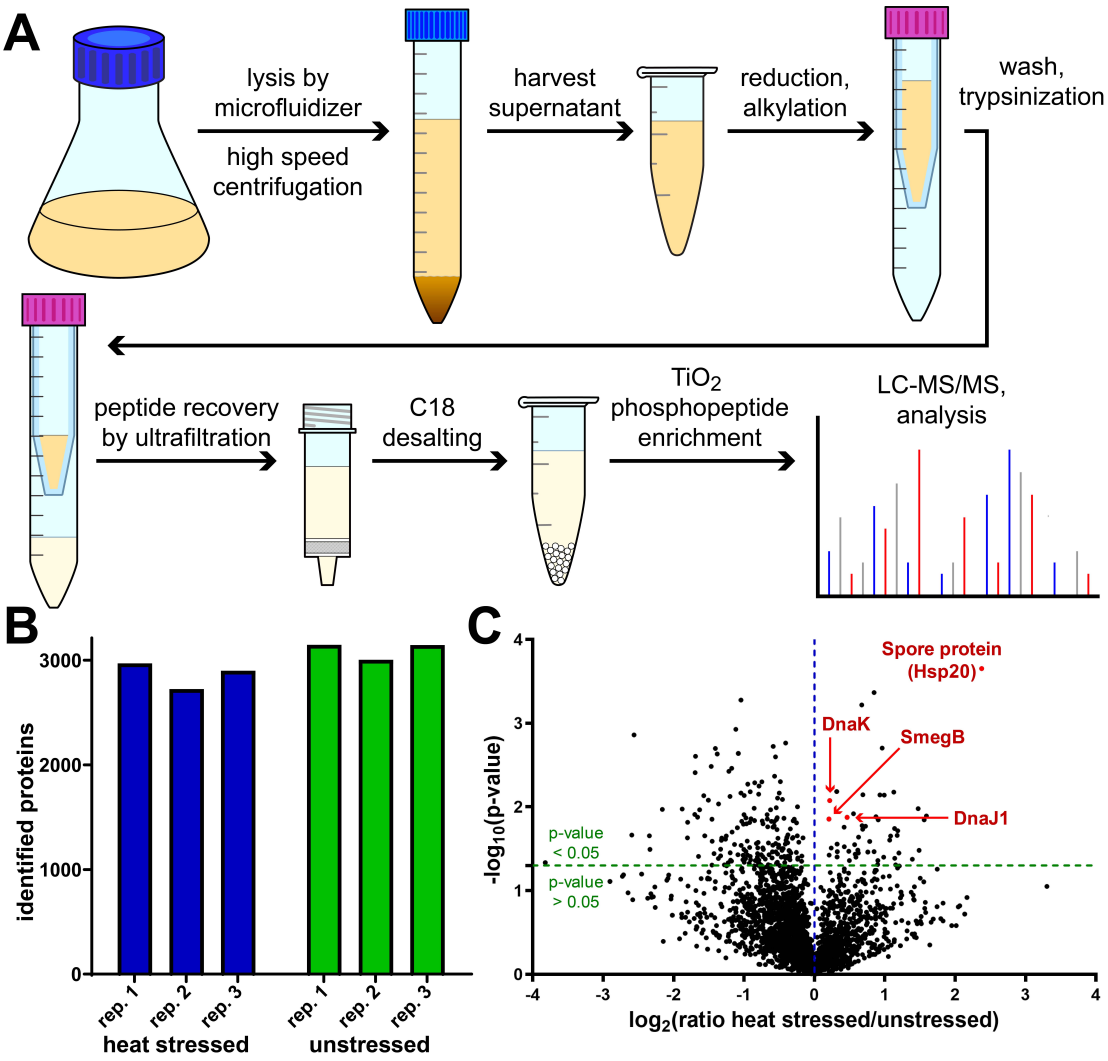



**Figure 2. Comparative analysis of mass spectrometry output. A**) Workflow for arginine phosphoproteomics in *Mycolicibacterium smegmatis*. **B**) Number of proteins obtained across biological replicates from cells grown in heat stress (50°C) and normal conditions (37°C). **C**) Volcano plot showing enrichment of proteins observed by LC-MS/MS in heat stress versus normal growth conditions. Plotted on the *x*-axis is the $\log_2$ of the ratio of average Sequest HT score of stressed to unstressed samples. The *y*-axis shows $-\log_{10}$ of p-values obtained by Student's t-test. The horizontal green line indicates a cut-off p-value = 0.05; the vertical blue line indicates score ratio = 1. Highlighted in red are notable chaperone proteins with stressed:unstressed ratio > 1.

**A**

| Protein | Accession | Gene | Position | Observations | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Unstressed | | | Stressed | | |
| | | | | 1 | 2 | 3 | 1 | 2 | 3 |
| Uncharacterized protein | A0QQU4 | MSMEG_0879 | R31 | • | • | • | • | • | • |
| Xanthine/uracil permeases family protein | A0QRZ5 | MSMEG_1293 | R475 | | • | | | | |
| Phosphate-specific transport system accessory protein PhoU1 | A0R4B7 | phoU | R149 | | • | • | • | • | • |
| Multifunctional 2-oxoglutarate metabolism enzyme | A0R2B1 | kgd | R433 | | • | | | | |
| Hemerythrin domain-containing protein | A0QQX2 | MSMEG_0907 | R34 | • | • | • | | • | • |
| Glycerol dehydratase large subunit | A0QSP1 | MSMEG_1547 | R326 | • | • | | • | • | • |

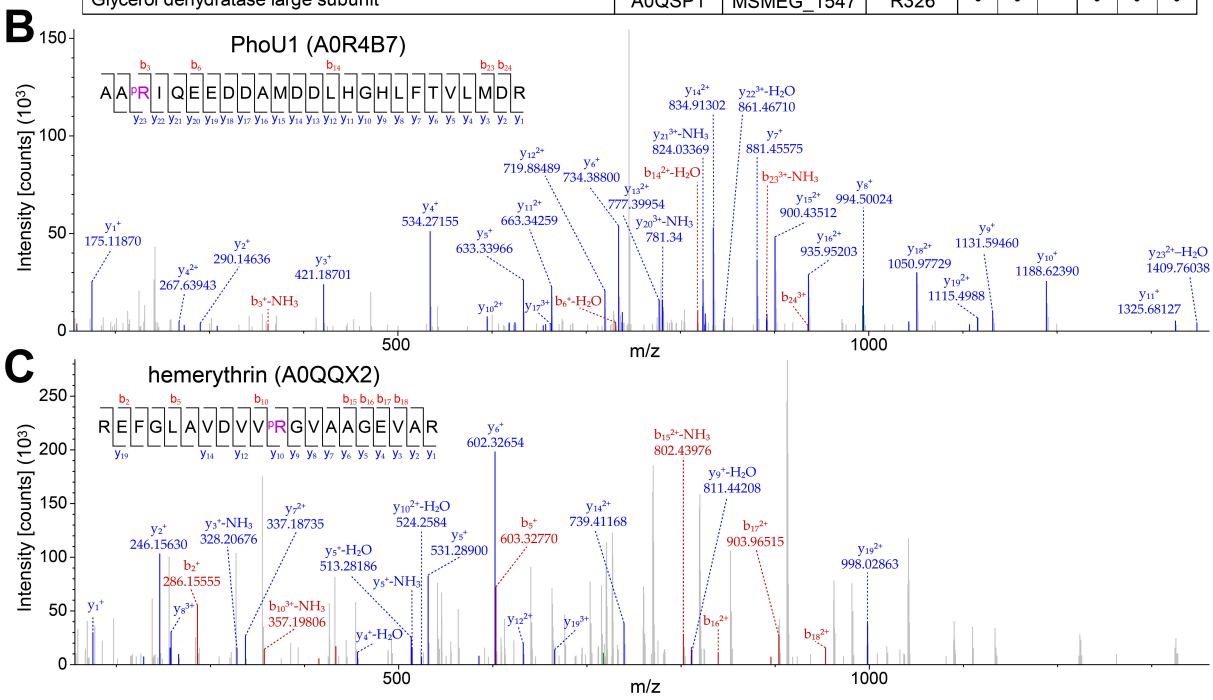



**Figure 3. Identification of arginine phosphosites in *Msm* proteome**. (**A**) Localized arginine phosphosites in six *Msm* proteins. Localization probability is reported as PhosphoRS/ptmRS score. Representative secondary fragmentation spectra show arginine-phosphorylated peptides from (**B**) Phou1 and (**C**) hemerythrin domain-containing protein. Additional spectra are shown in **Supplemental Figure S2.**

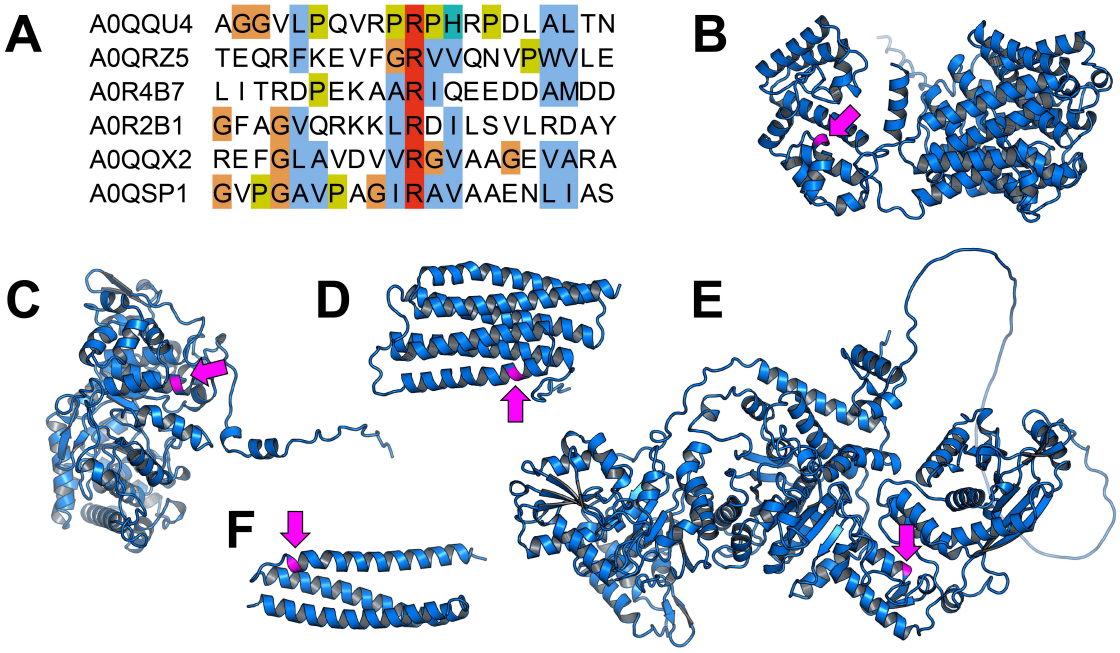



**Figure 4. Structural characteristics of phosphoarginine sites**. **(A)** Sequence alignment of 20 residues flanking the arginine phosphosites. Structural models of (**B**) Msmeg_1293, (**C**) Msmeg_1547, (**D**) PhoU1, (**E**) Kgd and (**F**) Msmeg_0907 are shown with the phosphorylated arginine colored magenta and marked with an arrow. All models were generated by AlphaFold2 (Jumper *et al.*, 2021).

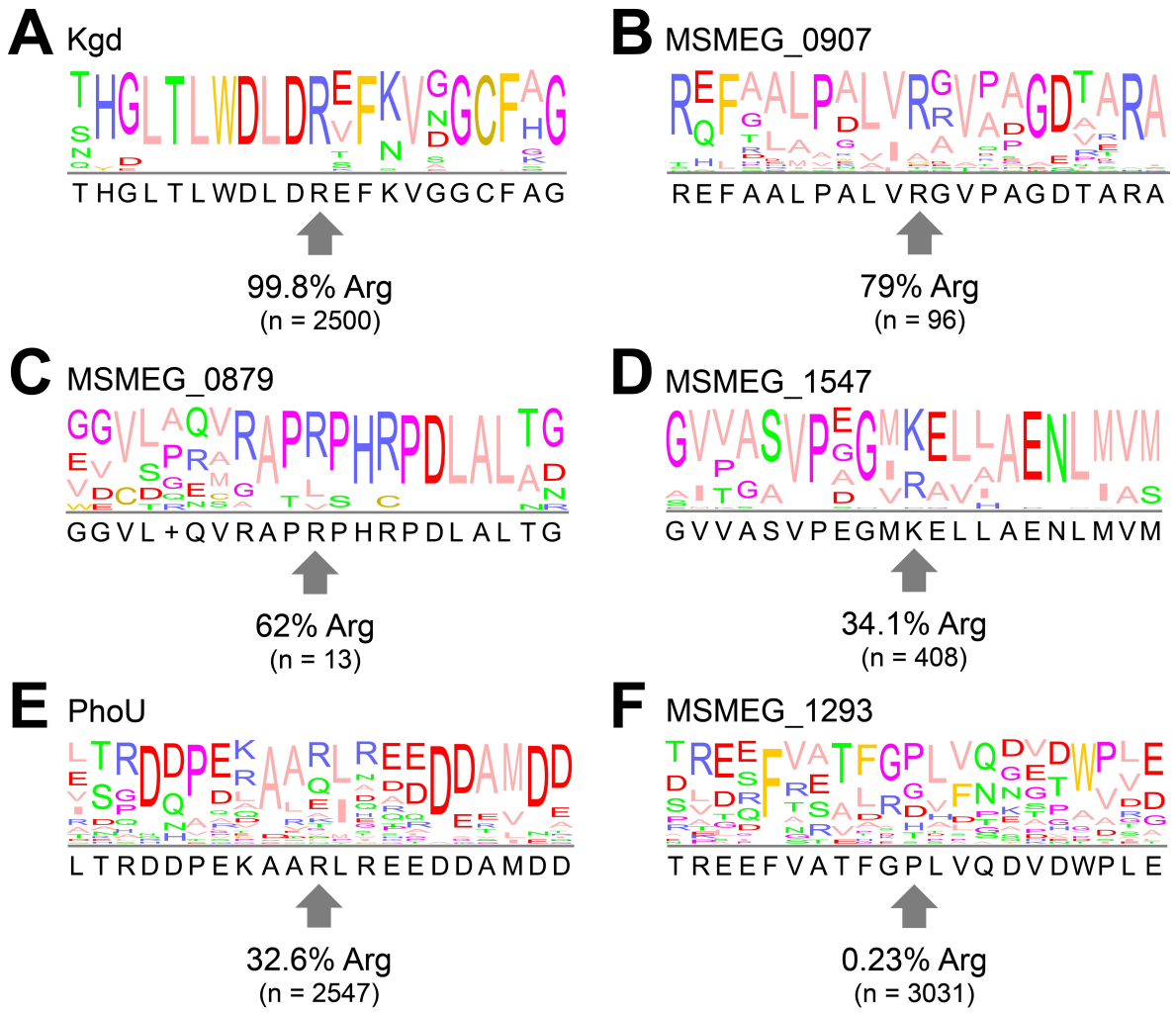
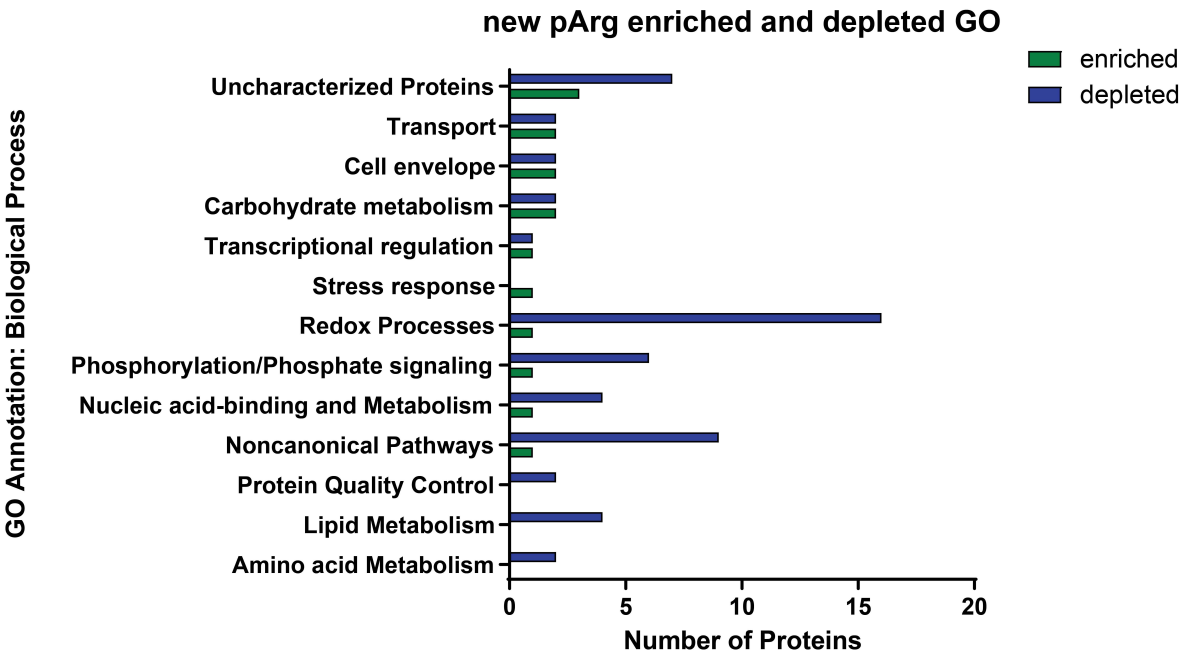
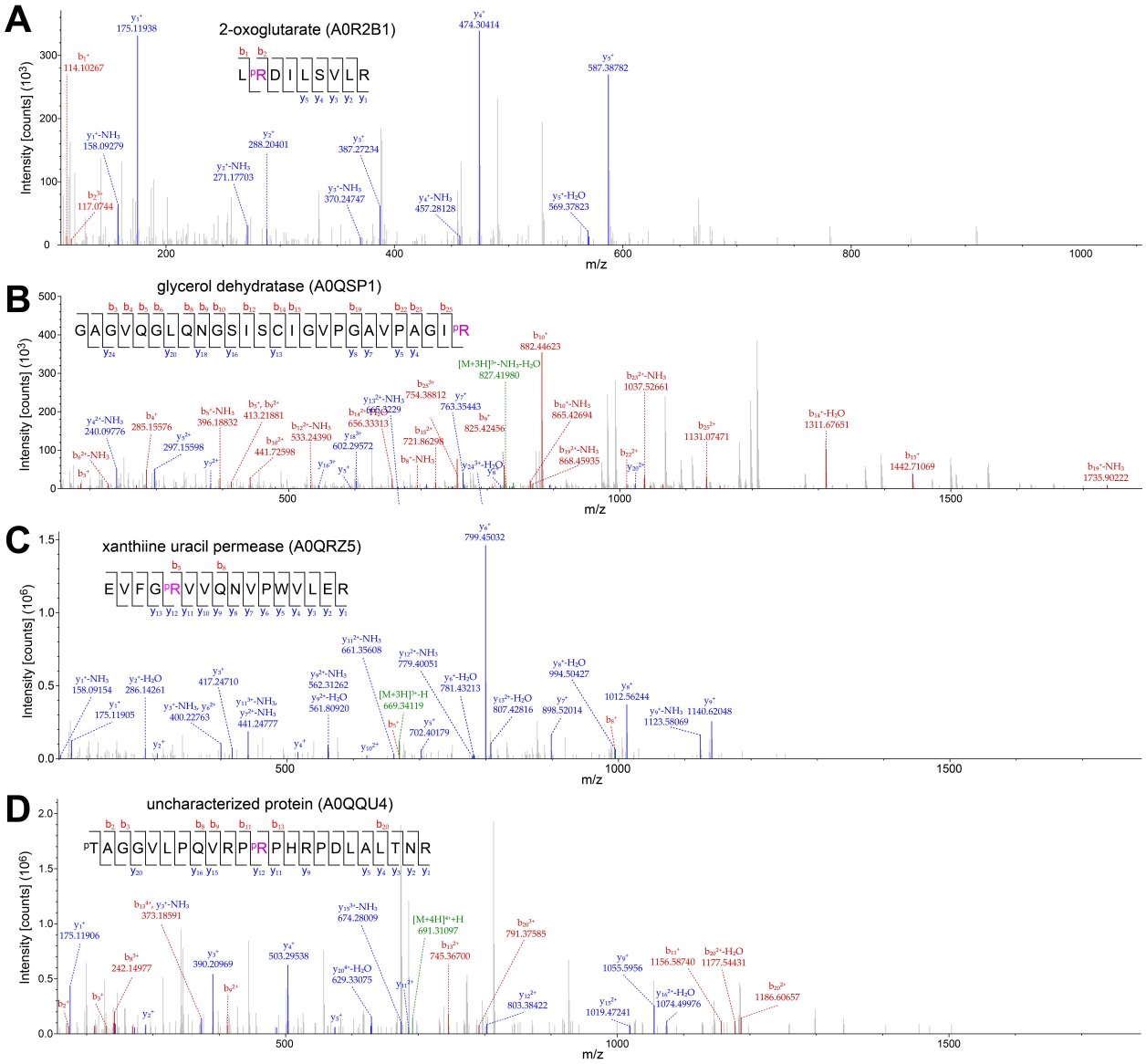


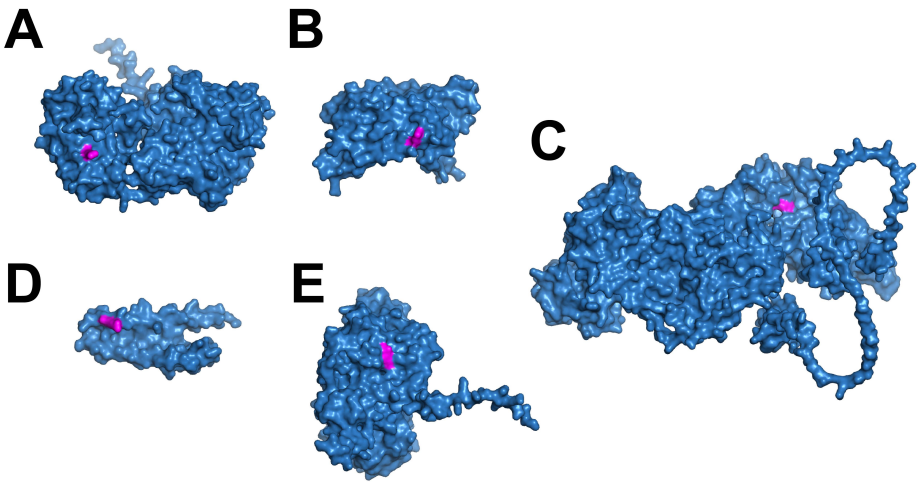

**Figure 5. Sequence conservation of phosphorylated arginine residues.** Sequence logos show conservation near phosphorylated arginines in (A) Kgd R433, (B) Msmeg_0907 R34, (C) Msmeg_0879 R31, (D) Msmeg_1547 R326, (E) PhoU1 R149, (F) Msmeg_1293 R475. Percentages indicates positional conservation of the phosphorylated arginine; “n” indicates the number of aligned orthologous sequences used to generate the sequence logo.




**Supplemental Figure 1. Gene ontology (GO) annotation of proteins strongly enriched or depleted upon heat shock.** Proteins with significantly ($p \leq 0.05$) and strongly different levels (two-fold or more) in heat shock and control conditions were sorted into GO annotation categories based on biological process.




**Supplemental Figure 2. Phosphosite secondary spectra.** Secondary fragmentation spectra showing arginine-phosphorylated peptides in (**A**) 2-oxoglutarate, (**B**) glycerol dehydratase, (**C**) xanthine uracyl permease, and (**D**) MSMEG_0879.




**Supplemental Figure 3. Solvent accessibility of phosphorylated arginines.** Surface representations of proteins (**A**) MSMEG_1293 (**B**) PhoU1 (**C**) Kgd (**D**) MSMEG_0907 (**E**) MSMEG_1547 based on AlphaFold2 prediction. Magenta patches represent specific arginine residues observed to be arginine-phosphorylated located on protein surface.