Published in final edited form as:

Mol Biosyst. 2010 February ; 6(2): 376–385. doi:10.1039/b916104j.

Expansion of the mycobacterial “PUPylome”†

Jeramie Watrous^a, Kristin Burns^b, Wei-Ting Liu^a, Anand Patel^c, Vivian Hook^d, Vineet Bafna^c, Clifton E. Barry 3rd^b, Steve Bark^d, and Pieter C. Dorrestein^{a,d,e}

Pieter C. Dorrestein: dorrestein@ucsd.edu

^a Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Ave., BSB 4090 MC 0636, La Jolla, CA 92093, USA. Fax: +1 858-822-0041; Tel: +1 858-534-6607

^b Tuberculosis Research Section, NIAID, National Institutes of Health, Bethesda, MD 20892, USA

^c Department of Computer Science, University of California, San Diego, La Jolla, CA 92093, USA

^d Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA

^e Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093, USA

Abstract

Selective degradation of cellular proteins offers an important mechanism to coordinate cellular processes including cell differentiation, defense, metabolic control, signal transduction and proliferation. While much is known about eukaryotic ubiquitination, we know little about the recently discovered ubiquitin-like protein in prokaryotes (PUP). Through expression of His₇ tagged PUP and exploitation of the characteristic +243 Da mass shift attributed to trypsinized PUPylated peptides, a global pull-down of protein targets for PUPylation in *Mycobacterium smegmatis* revealed 103 candidate PUPylation targets and 52 confirmed targets. Similar to eukaryotic ubiquitination, further analysis of these targets revealed neither primary sequence nor secondary structure homology at the point of attachment. Pathways containing PUPylated proteins include many central to rapid cell growth, such as glycolysis, gluconeogenesis, amino acid and mycolic acid metabolism and biosynthesis, as well as translation. Seventeen of the 29 nitrosylated protein targets previously identified in *Mycobacterium tuberculosis* were also identified as PUPylation candidates indicating a connection between PUP-mediated remodeling of critical metabolic pathways and the mycobacterial response to exogenous stress.

Introduction

Cellular pathways involved in determining the fate of essential proteins through post-translational events have become an increasingly important area of study.^{1–5} Of these modifications, the understanding of eukaryotic ubiquitination has shown to be especially valuable. With the ability to mark specific proteins for proteasomal degradation, this pathway has been shown to play a critical role in the cell cycle, cellular metabolism, cell signaling and immune response.^{6,7} Disruption of this pathway can lead to many human diseases such as

†Electronic supplementary information (ESI) available: Summary results for all proteins found in this study (ST1–ST4), a comparison of nitrosylated and PUPylated proteins (ST5), manual annotations of mass spectra for all confirmed PUPylated proteins (SF1–SF48), sequence alignment of all PUPylated proteins centered at the site of modification (SF49) as well as full length (SF50), individual sequence alignments of PUPylated proteins showing extent of evolutionary sequence conservation at site of modification (SF51–SF102), structural annotations showing site of PUPylation on homologous proteins (SF103–SF123) and pathway annotations showing locations of PUPylated and candidate proteins on Msm pathway maps (SF124–SF136).

Correspondence to: Pieter C. Dorrestein, dorrestein@ucsd.edu.

inflammation, cancer, Parkinson's disease and Alzheimer's disease making it a very attractive drug target.^{3,8} While the eukaryotic ubiquitin–proteasome degradation pathway was discovered in the late 1970's,⁹ it was only recently that a ubiquitin-like protein was identified in prokaryotes.^{1,4} This new prokaryotic ubiquitin protein (PUP) has now been characterized from the Actinobacterium *Mycobacterium tuberculosis* (Mtb) and its avirulent relative *Mycobacterium smegmatis* (Msm).

Similar to ubiquitin, PUP has been implicated in proteasomal degradation.^{1,4} The PUP dependent pathway was discovered *via* two independent approaches. Pearce *et al.* identified the prokaryotic ubiquitin-like protein from Mtb *via* an affinity pull-down with the mycobacterium proteasomal AAA+-ATPase (Mpa). From previous work, these authors had already identified that FabD, the malonyl acyltransferase on the fatty acid biosynthetic pathway, and PanB (ketopantoate hydroxymethyltransferase), the bifunctional enzyme involved in the biosynthesis of coenzyme A, were substrates for proteasomal degradation.¹⁰ They therefore attached a His₆ tag to FabD and determined that the protein was PUPylated and that PUPylation was a marker for FabD proteolysis. In our study, the PUP was discovered using comparative genomic analysis.¹ Comparative genomics takes advantage of the fact that many prokaryotic organisms cluster genes on the chromosome for specific functions. In this case, a smaller protein of ~7 kDa was consistently clustered with putative proteasomal subunits and a homolog of the ATPase Mpa. These gene clusters were not only found in Mtb and Msm, but also in distantly related Actinobacteria such as the plant microbe *Frankia*, the marine Actinobacterium *Salinispora tropica*, as well as Streptomyces such as *Streptomyces coelicolor* and *Streptomyces avermitilis*.^{1,2}

The sequence of PUP bares little resemblance to that of ubiquitin, but one can observe some similarities.^{1,4} The C-terminal Gly-Gly of ubiquitin is activated by ATP *via* a protein cascade involving E1, E2 and E3 type proteins before it labels the target protein, often resulting in proteasomal degradation (Fig. 1A). In most Actinobacterium, PUP also has a C-terminal Gly-Gly, but is then followed by a glutamine (in some Actinobacteria a glutamate). Peptides originating from a trypsinized ubiquitinated protein display a characteristic 114.1 Da mass shift due to the Gly-Gly remnant still bound to the target lysines. Using a tagged ubiquitin purification strategy and LC-MS/MS, Peng *et al.* used this mass shift to mine proteomic data sets and identified 1075 candidate ubiquitinated proteins containing 110 ubiquitination sites from 72 different proteins from humans.⁵ In the case of PUPylated proteins, we (and others) have shown that a trypsin digest of PUPylated substrates leaves a GGQ modification on the target lysine.^{1,4} In addition, the Q is deamidated giving rise to a 243.081 Da addition to the target lysine. While the entirety of the PUPylation mechanism has yet to be elucidated, recent findings by Striebel *et al.* have shown that conjugation of PUP to its target protein occurs through a two step process in Mtb.¹¹ PUP is initially deamidated by the PUP deamidase Dop followed by conjugation to the target protein *via* the carboxylate amine/ammonia ligase-like protein PafA (Fig. 1B). There is little to no evidence regarding the nature of substrates that are PUPylated; however, studies have shown that the rate of degradation of PUPylated proteins is significantly reduced in proteasome mutants, which indicates that PUPylated proteins are likely degraded by the proteasome.^{1,4}

In the present work we set out to address two objectives: to identify other PUPylated substrates in an effort to elucidate the pathways that are PUPylated and to establish a sequence motif for PUPylation. Thus far two PUPylation sites have been characterized and two additional proteins, superoxide dismutase and PanB, have been suggested to be substrates for PUPylation.^{1,4,10} His₇ tagged PUP in combination with LC-MS/MS and database searches to identify PUPylated proteins provided us with additional insight into the nature and function of PUPylation. Detailed analysis of the PUPylome enabled us to implicate this mechanism of controlled protein degradation in Mtb cellular defense. Previous work has shown that the proteasome in Mtb is

essential for cell proliferation in the host organism and therefore PUPylation is likely to play a role in the cell viability of Mtb and could possibly be an important target for therapeutic drugs.

Experimental

Strains and molecular biology

Growth of strains, primer construction, cloning expression, and purification were conducted according to the previously established protocol by Burns *et al.*¹ Briefly, primers spolyGf: TACCAAGCTTGGAGCCCCTGCGCGGGAG and smeghis2: CATATGATGATGATGATGATGATGCATCGCTGCCTCCTGCAAAAGTC were used to amplify the region about 200 bp upstream of the *pup* gene from Msm and to introduce a 7×-histidine tag at the amino terminus. Primers pupfor: AGGAGGCACATATGGCTCAGGAG and puprev2: GATTATCGCGGATCCTCACTGGCC were used to amplify the *pup* gene. PCR products were digested with HindIII/NdeI and Nde/BamHI (New England Biolabs), respectively. The two digested PCR products were ligated into the polyG vector, digested with HindIII/BamHI and dephosphorylated. The ligated DNA was transformed into *E. coli* One Shot TOP10 cells (Invitrogen), and successful clones (hisPup.polyG) were sequenced (Macrogen). Clones were electroporated into competent Msm MC²-155 following established protocols.¹² Transformed cells were grown in 7H9 media to an OD₆₅₀ of 0.8–1 after which they were harvested, lysed by sonication, and purified by Ni-NTA chromatography under denaturing and native conditions as suggested in the Qiagen manual. Denaturing purification included 8 M urea and used a pH gradient to elute the bound 7×-histidine tag while elution under native conditions included an imidazole containing elution buffer. The resulting purification by SDS-PAGE can be seen in Fig. 2.

Preparation of tryptic peptides

The lyophilized protein samples were dissolved in 25 mM Tris buffer, pH 7.8, containing 10% glycerol to give a final concentration of approximately 2 µg µL⁻¹. A portion of the samples were subjected to a standard trypsin digest using a Sigma-Aldrich Trypsin Singles Kit (Cat No: T7575) where 50 µL (equivalent to 100 µg of protein) of each sample were added separately to individual microfuge tubes containing 1 µg of lyophilized trypsin and 1 µL of trypsin solubilizing solution provided in the kit. To each tube 49 µL of lysis buffer, which is also supplied in the kit, was added and the samples were allowed to digest for 21 h at room temperature and were quenched with 100 µL of 10% formic acid to yield a final concentration of approximately 0.5 µg µL⁻¹. The samples were then subjected to mass spectrometric analysis on three different instruments in order to yield high confidence overlapping data.

Tandem-MS analysis

The samples were initially analyzed by direct infusion using an Advion Nanomate 100 robot coupled with a 6.42T LTQ-FT-ICR-MS (ThermoFinnigan). Each sample was loaded onto a C₁₈ ZipTip (equilibrated with 200 µL of 5% ACN in water with 0.1% acetic acid) by pushing through 50 µL of the digested sample solution followed by washing with 200 µL of 5% ACN in water with 0.1% acetic acid. The peptides were then eluted by pushing through 20 µL each of a series of elution solutions ranging from 15% ACN in water with 0.1% acetic acid to 85% ACN in water with 0.1% acetic acid in 10% increments directly into a 96 well plate. Each sample was introduced into the LTQ-FT-ICR-MS *via* the nanomate with a starting voltage of 1.45 kV and an initial back pressure of 0.40 bar. The voltage and back pressure were adjusted slightly during the injection to maintain a suitable current. Data were initially collected from 400 to 2000 *m/z* on the FT-ICR-MS at 100000 resolution followed by data dependent analysis on the LTQ. The first scan event was broadband MS1 followed by five subsequent scans dependent on the data collected in the previous scan event. The dynamic exclusion was set to

a repeat count of 1 spectra and a repeat duration of 30 s. The exclusion list was initially set at 200 ions with an exclusion duration of 80 s, but was later changed to 500 ions with an exclusion duration of 240 s in an effort to gather data for lower abundant ions. Data were collected for each sample for approximately 5 min using ChipsoftManager (Advion; version 7.2.0.530) to control the nanomate, Tune Plus 1.0 (Thermo Scientific) to tune the mass spectrometer and Xcalibur (Thermo Scientific) to control the LTQ-FT-ICR-MS.

In order to obtain more complete coverage of the digest, the samples were subjected to the same trypsin digest described above with the exception that the initial sample was diluted 1 : 1 with 25 mM Tris pH 7.8 containing 10% glycerol to yield a 1 : 50 trypsin to protein digest as opposed to a 1 : 100 digest used previously. Once quenched, the samples were lyophilized and re-dissolved in 25 μL of 5% ACN in water containing 0.1% acetic acid yielding a final concentration of approximately 1 $\mu\text{g } \mu\text{L}^{-1}$. These samples were analyzed on three different mass spectrometers.

The first set of analysis was performed on an Agilent 1200 quaternary pump HPLC coupled to a Thermo Scientific LTQ XL. The peptides were separated on a 10 cm $\times$ 100 μm $\times$ 5 μm C₁₈ packed capillary column using a mobile phase composition of (A) 0.1% FA in 5% ACN and 95% water and (B) 0.1% FA in 90% ACN and 10% water running the following gradient: isocratic from initial to 21 min at 100% A with increasing flow rate from 7 $\mu\text{L min}^{-1}$ to 150 $\mu\text{L min}^{-1}$ followed by a linear gradient from 21 (100% A) to 84 min (20% A). The sample compartment was kept at 5 °C while the column temperature was at ambient and the injection volume was set to 10 μL . Data dependent analysis was performed with the initial scan event being broadband MS1 from 400 to 2000 m/z followed by five additional scan events dependent on the preceding scan. The dynamic exclusion was set to a repeat count of 1 spectra and a repeat duration of 30 s. The exclusion list was initially set at 100 ions with an exclusion duration of 180 s. Tune Plus 1.0 (Thermo Scientific) was used to tune the mass spectrometer while Xcalibur software (Thermo Scientific) was used for the analysis.

Further analysis using the samples from the 1 : 50 trypsin digest was performed on an Agilent 1200 nano-flow HPLC coupled to an Agilent 6330 Ion Trap. The peptides were separated on a Zorbax 300SB-C₁₈ 5 μm (150 mm $\times$ 75 μm) chip cube column using a mobile phase composition of (A) 0.25% FA in water and (B) 0.25% FA in ACN flowing at 0.40 $\mu\text{L min}^{-1}$ running the following gradient: initial to 50 min a linear gradient from 97% A to 55% A and from 50 to 60 min a linear gradient from 55% A to 5% A. The sample compartment was kept at 5 °C while the column temperature was at ambient and the injection volume was set to 1 μL . Data dependent analysis was performed from 400 to 1800 m/z with the initial scan event being broadband MS1 followed by 4 additional scan events dependent on the preceding scan. The active exclusion was set to 1 spectra with a repeat time of 30 s. Chemstation for LC 3D Systems software (Agilent; version B.01.03 SR1) was used for the analysis.

Lastly, analysis using the 1 : 50 tryptic digest was performed on a ThermoFinnigan Surveyor coupled to the previously used LTQ-FT-ICR-MS. The peptides were separated on a 100 μm $\times$ 5 μm C₁₈ packed capillary column using a mobile phase composition of (A) 0.1% AcOH in water and (B) 0.1% AcOH in ACN running the following gradient: isocratic from initial to 15 min at 95% A at 250 $\mu\text{L min}^{-1}$ followed by a linear gradient from 15 (95% A at 250 $\mu\text{L min}^{-1}$) to 120 min (30% A at 500 $\mu\text{L min}^{-1}$). The sample compartment was kept at 4 °C while the column temperature was at ambient and the injection volume was set to 10 μL . Data were initially collected from 400 to 2000 m/z on the FT-ICR-MS at 12 500 resolution followed by data dependent analysis on the LTQ. The first scan event was broadband MS1 followed by five subsequent scans dependent on the data collected in the previous scan event. The dynamic exclusion was set to a repeat count of 1 spectra and a repeat duration of 30 s. The exclusion list was initially set at 100 ions with an exclusion duration of 180 s. Tune Plus 1.0 (Thermo

Scientific) was used to tune the mass spectrometer while Xcalibur (Thermo Scientific) was used for the analysis.

Data analysis

The resulting .RAW (from the Xcalibur software) and .YEP (from the Agilent software) data files were converted into .mzXML format using ReadW (tools.proteomecenter.org/TPP.php) and CompassXport (www.brukerdaltonics.com), respectively. The data were analyzed using InSpecT (proteomics2.ucsd.edu/LiveSearch/index.jsp), which was developed at the UCSD NIH NCRR center for computational mass spectrometry, to look for post-translational modifications using peptide sequence tags from the tandem mass spectrometry data to filter the database.¹³ In addition, all spectra were clustered using the default algorithm and the LTQ low pass filter of MS Cluster (proteomics2.ucsd.edu/LiveSearch/index.jsp) resulting in 47–54% of spectra falling into clusters with the remaining spectra filtered out as noise. Clustering averages common mass spectra which allows for enrichment of common peaks due to reduction in random noise.¹⁴ Clustering was done to maximize the mining of the data with InSpecT as it can increase the spectral annotation by 10–25%. Previous work has shown that a trypsin digested PUPylated peptide displays a +243.081 Da mass shift due to the addition of a deamidated QGG motif.¹ InSpecT was therefore programmed to search for a +243 modification with a parent mass tolerance of 2 Da and b/y ion mass tolerances of 0.5 Da. The data were searched against the proteomic database for Msm strain MC²-155 (www.NCBI.com), a database of common contaminants (such as human keratin and porcine trypsin) and a reverse shuffled database for Msm strain MC²-155 to identify spurious hits. The output files for every sample from every instrument were compiled into one master data set. A strict 1% false positive cutoff (sorted by the MQ score provided by InSpecT) was used for all non-PUPylated peptides while the spectra for all PUPylated peptides were manually annotated except for modified peptides found only in the clustered data sets as these spectra represent an average of multiple spectra. The resulting data set was summarized using InSpecT into html format, which included annotated spectra, using an additional cutoff of *p*-value < 0.05 resulting in a final false discovery rate cutoff of 0.95%. For all the PUPylation sites discovered in the InSpecT search for the unclustered data set, the annotated spectra are provided in the ESI[†].

Results and discussion

Identification of candidate PUPylated proteins

The first major goal of the present work was to expand the list of protein targets for PUPylation. For this purpose, the Msm PUP was tagged at the N-terminus with a His₇ tag and expressed in its native host. As a control, PUP was expressed without the His₇ tag, which is here forth referred to as the non-tagged sample. Cell lysates were purified using Ni-NTA affinity chromatography under both denaturing and native conditions in order to identify proteins targeted for PUPylation and proteins non-covalently interacting with PUPylated proteins, respectively. Fig. 2 shows the eluted proteins from these samples. As expected, purification of the His₇ tagged PUP sample under both denaturing and native conditions results in a ladder of proteins. Following purification, the samples were trypsinized and analyzed by LC-MS/MS. The resulting spectra were clustered using MS Cluster to reduce random noise and subsequent analysis of both clustered and unclustered data sets using the open source proteomics program InSpecT revealed a total of 245 proteins, of which 52 proteins were confirmed to have a lysine

[†]Electronic supplementary information (ESI) available: Summary results for all proteins found in this study (ST1–ST4), a comparison of nitrosylated and PUPylated proteins (ST5), manual annotations of mass spectra for all confirmed PUPylated proteins (SF1–SF48), sequence alignment of all PUPylated proteins centered at the site of modification (SF49) as well as full length (SF50), individual sequence alignments of PUPylated proteins showing extent of evolutionary sequence conservation at site of modification (SF51–SF102), structural annotations showing site of PUPylation on homologous proteins (SF103–SF123) and pathway annotations showing locations of PUPylated and candidate proteins on Msm pathway maps (SF124–SF136).

modified with the characteristic +243 Da modification corresponding to PUPylation (Table 1). After manual inspection of these 52 proteins, 76% of peptides showing the +243 Da mass shift from the unclustered data set also displayed a unique MS2 fragmentation pattern where the characteristic loss of GG and Q*GG could be observed, further solidifying their identities (Q* refers to deamidated glutamine). Some examples of this characteristic GG and Q*GG loss are shown in Fig. 3 (and Fig. SF1–SF48, ESI[†]), and could become a training marker to confirm additional PUPylated proteins in future proteomic data sets of organisms that contain a PUP protein. It should be noted that these fragment ions are often some of the most intense ions within the spectrum indicating that the amides of the K–Q* terminal PUP linkage and the Q*–G are activated in an analogous manner to amides on the N-terminal side of proline and the C-terminal side of glutamates and aspartates.⁵

The remaining 193 proteins were subdivided into candidate and non-candidate PUPylation targets based on the distribution of their peptides between the His₇-tagged and non-tagged control samples. Initial analysis of the peptide distribution of the 52 confirmed PUPylated proteins showed that as more peptides were seen in the non-tagged control sample, the prevalence of PUPylated proteins decreased. Using this trend, we utilized spectral counting to correlate the enrichment of tryptic peptides in the His₇-PUP and non-tagged control samples with the probability that a protein is PUPylated. In our spectral counting, we combined non-clustered spectra from all the LC-MS/MS sample runs and plotted the distribution of spectral counts for each protein between the non-tagged control, His₇-PUP sample purified under native conditions and the His₇-PUP sample purified under denaturing conditions in an xyz-plot (Fig. 4 and Tables ST1–ST4 (ESI[†])). By excluding any protein containing more than 44.6% of its peptides in the non-tagged control from consideration as a PUP target (equivalent to the highest observed confirmed PUPylated protein), 103 of the 193 remaining proteins were found to be candidates for PUPylation (see Table ST2).[†] In addition, 12 proteins were found only in the His₇-PUP sample purified under native conditions and therefore we have annotated these candidate proteins to non-covalently interact with PUPylated proteins (see Table ST3).[†]

Since quantification with spectral counting has inherent limitations with low spectral counts,^{15,16} we set a cutoff for the minimum number of spectra for a candidate protein to be included into the analysis at 10. While this stringent cutoff of 10 spectra per protein eliminated some high probability candidates such as nitrogen regulatory protein P-II, cold shock DNA binding domain, and the 30S ribosomal protein S7, we felt it important to provide more stringent cutoff of the data for further interpretation.

Unlike the human ubiquitinome where multiple ubiquitination sites are often found for a given protein substrate, only one of the 52 confirmed PUPylated proteins, [Mn]superoxide dismutase, had more than one PUPylation site. In addition, no evidence was obtained to support poly-PUPylation despite having 1661 spectra covering the PUP protein. This suggests that PUP does not form polymeric chains with itself as seen in the eukaryotic ubiquitin pathway, which is in agreement with previous observations where comparative analysis of PUPylated and non-PUPylated FabD and *myo*-inositol-phosphate synthase showed only a single new band from SDS-PAGE and Western blot analysis, respectively.^{1,4,11}

PUPylated proteins lack a primary sequence recognition motif

To determine if the PUPylation sites contained a specific recognition motif in the primary sequence we aligned the modified lysines of the 52 confirmed PUPylated proteins with ± 10 residues in each direction of the site of attachment (see Fig. SF49, ESI[†]). While other post-translational modifications such as glycosylations, phosphorylations, and phosphopantetheinylations display characteristic sequence motifs at the site of modification, no apparent recognition motif could be found for PUPylation. In addition, a hydrophobicity plot failed to identify any conserved hydrophobic or hydrophilic patches around the target

lysine (see Fig. SF49).[†] An alignment of the complete primary sequence of all 52 confirmed proteins using the ClustalW algorithm also failed to show a common off-site recognition sequence (see Fig. SF50).[†] To determine if the modification sites were conserved amongst the phylum, homologs for all 52 PUPylated proteins were found using NCBI protein BLAST. Thirteen different species from the Actinobacterial phylum were chosen and aligned *via* ClustalW algorithms when available (see Fig. SF51–SF102).[†] The alignments showed that while the majority of the primary sequences of the PUPylated proteins were fairly conserved across the phylum, the PUPylation site was not, with only 25% conservation of the lysine residue. Even among *Mycobacterium* only 65% of the proteins contained a conserved lysine at the site of PUPylation. Unlike eukaryotic ubiquitination, this lack of conservation at the target site indicates that PUPylated proteins, target residues, and/or sequence motifs may vary from species to species even within the same phylum.

The lack of a primary sequence motif suggested a possible 3-dimensional component to site recognition. When the locations of the modified lysines of several PUPylated proteins were plotted on all the available crystal structures of homologous proteins they were consistently on the periphery of the target proteins as well as on peripheral subunits for multimeric proteins (see Fig. SF103–SF123, ESI[†]). The modification site was usually located on either an alpha helix or an unstructured protein region with no observed modification sites on beta sheets. Homologs for the E1, E2 and E3 complexes, which facilitate the interaction of ubiquitin with proteins marked for degradation in eukaryotes, have yet to be identified in prokaryotes. These results suggest that the mechanism by which these proteins are marked for proteasomal degradation may prove complex.

Pathways regulated through PUPylation

To provide insight into the pathways controlled by PUPylation, all 155 confirmed and candidate proteins were searched using NCBI BLAST to obtain their assigned enzyme codes, if available. Using the bioinformatics database Kegg Atlas (www.genome.jp/kegg), the 103 proteins with an enzyme code were mapped onto pathways from *Msm* MC²-155 (see Fig. SF124–SF136, ESI[†]). This analysis showed an abundance of PUPylation in metabolic pathways such as glycolysis/gluconeogenesis, citrate cycle, fatty acid metabolism, purine metabolism and pyruvate metabolism. Biosynthetic pathways also seem to be highly regulated by PUPylation, especially the biosynthesis of valine-leucine-isoleucine and fatty acids. One of the most striking examples of pathway regulation can be seen in the cluster of genes primarily responsible for mycolic acid biosynthesis (Fig. 5).¹⁷ We were able to confirm PUPylation for 5 of the proteins on this particular gene cluster. *Mycobacteria* have an unusually large number of enzymes involved in fatty acid biosynthesis (roughly 5 times that of *E. coli*) and must balance *de novo* synthesis of normal length fatty acyl components of the plasma membrane with the considerably longer mycolic acids that are components of the mycobacterial outer cell envelope.¹⁸ Mycolic acids represent up to one-third of the dry weight of mycobacterial cells and their synthesis is one of the most energetically costly activities undertaken by the bacilli.¹⁹ Mycolic acids and fatty acids originate from a common pool of precursors and very little is known about the regulation of these pathways. PUPylation may play an important role in allowing the cell to rapidly switch off metabolic flux of lipids into this critical pathway. Consistent with this apparent role in regulation of some of the most metabolically important pathways in the mycobacterial cell, the ribosome is also a target of PUPylation with 3 confirmed and 4 candidate PUPylated ribosomal subunits found in this study.

Another observation was the PUPylation of the *Msm* Mpa AAA+-ATPase homolog (MSMEG_3902). The AAA+-ATPase is essential in the unfolding and translocation of target proteins into the proteasome core.²⁰ The fact that it is targeted for PUPylation indicates a possible homeostatic mechanism for controlling the rate of proteasomal degradation within the

cell. This finding is in agreement with the observed turnover of the AAA+-ATPase.¹⁰ It has also been shown that deletion of the AAA+-ATPase in Mtb decreases growth rate and virulence, but the cells persist within the host.²¹ Thus careful control of PUPylation, and this homeostatic mechanism for limiting the extent of proteolysis, may constitute an important overall component of the ability of Mtb to respond to changing intracellular conditions with an appropriate rate of growth.

Role of PUPylation in nitric oxide (NO) defense

One of the major host immune responses against bacterial pathogens like Mtb are reactive nitrogen intermediates (RNI's), catalyzed by the enzymes iNOS and NOS2.^{22–24} These RNI's can inflict damage to the pathogen *via* nitric oxide, which can nitrosylate essential proteins rendering them inactive, or react with superoxide (a byproduct from bacterial metabolism) to form peroxynitrite, which inflicts oxidative damage on the cell.²² To combat the host immune response, the Mtb cell surface receptors, DosS and DosT, have been shown to bind NO and trigger a kinase signaling cascade resulting in decreased cellular metabolism and stunted cell growth and proliferation allowing Mtb to enter a non-replicating form where it can survive in a human host for years.^{24–26} A global survey of proteins in Mtb targeted for nitrosylation identified 29 proteins,²⁴ 17 of which we have either confirmed or identified as candidate PUPylation targets in this study in Msm (see Table ST5, ESI[†]). The correlation between nitrosylated and PUPylated proteins could indicate that PUPylation plays a more complex role in turning over nitrosylated, and therefore inactivated proteins *via* proteasomal degradation.²⁷ Our data further support this hypothesis with the finding that alkylhydroperoxide reductase C, 2-oxoglutarate dehydrogenase and dihydrolipoamide dehydrogenase were found to be candidates for PUPylation. These proteins constitute 3 of the 4 NADH dependent peroxidase and peroxynitrite reductase proteins encoded by Mtb, which have been shown to be essential in Mtb antioxidant defense and pathogenesis.^{28,29}

Conclusions

In this study, we have identified 52 proteins targeted for PUPylation along with an additional 103 possible candidates. Despite this large sample, neither a primary sequence nor secondary structure recognition motif could be identified. However, based on the identified PUPylated proteins, it is clear that essential cellular growth pathways such as the valine-leucine-isoleucine and mycolic acid biosynthetic pathway as well as the ribosome are targeted. In addition, the PUPylation pathway appears to be self-regulated by its ability to PUPylate the AAA+-ATPase, a protein known to not only be turned over but also a protein required for the degradation of substrates. Finally, the identification of a significant overlap between the nitrosoproteome and the PUPylome suggests an important possible function of the PUPylation system. PUPylation may have arisen in this relatively slowly replicating group of organisms as a rapid response mechanism for modulating growth rate in response to exogenous stress. If this hypothesis proves true in Mtb then PUPylation may play a central role in the phenomenon of latent tuberculosis infections that afflict one-third of the global population.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported (in part) by the Intramural Research Division of the NIAID, the Skaggs School of Pharmacy and Pharmaceutical Sciences (P.D.), the NIH Molecular Biophysics Training Program GM08326 (J.W.), NIH Grant P01 HL58120 (V.H.), NIH Grant R01 DA04271 (V.H.), and NIDA and NIH Grant 5K01DA23065 (S.B.).

References

1. Burns KE, Liu WT, Boshoff HI, Dorrestein PC, Barry CE. J Biol Chem 2009;284:3069–3075. [PubMed: 19028679]
2. Iyer LM, Burroughs AM, Aravind L. Biol Direct 2008;3:45. [PubMed: 18980670]
3. Kisselev AF, Goldberg AL. Chem Biol 2001;8:739–758. [PubMed: 11514224]
4. Pearce MJ, Mintseris J, Ferreyra J, Gygi SP, Darwin HK. Science 2008;322:1104–1107. [PubMed: 18832610]
5. Peng J, Schwartz D, Elias JE, Thoreen CC, Cheng D, Marsischky G, Roelofs J, Finley D, Gygi SP. Nat Biotechnol 2003;21:921–926. [PubMed: 12872131]
6. Matthews W, Driscoll J, Tanaka K, Ichihara A, Goldberg AL. Proc Natl Acad Sci U S A 1989;86:2597–2601. [PubMed: 2539595]
7. Sassetti CM, Boyd DH, Rubin EJ. Mol Microbiol 2003;48:77–84. [PubMed: 12657046]
8. Hol E, Vanleeuwen F, Fischer D. Trends Mol Med 2005;11:488–495. [PubMed: 16213790]
9. Etlinger JD, Goldberg AL. Proc Natl Acad Sci U S A 1977;74:54–58. [PubMed: 264694]
10. Pearce MJ, Arora P, Festa RA, Butler-Wu SM, Gokhale RS, Darwin HK. EMBO J 2006;25:5423–5432. [PubMed: 17082771]
11. Striebel F, Imkamp F, Sutter M, Steiner M, Mamedov A, Weber-Ban E. Nat Struct Mol Biol 2009;16:647–651. [PubMed: 19448618]
12. Jacobs WR, Kalpana GV, Cirillo JD, Pascopella L, Snapper SB, Udani RA, Jones W, Barletta RG, Bloom BR. Methods Enzymol 1991;204:537–555. [PubMed: 1658570]
13. Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, Bafna V. Anal Chem 2005;77:4626–4639. [PubMed: 16013882]
14. Frank AM, Bandeira N, Shen Z, Tanner S, Briggs SP, Smith RD, Pevzner PA. J Proteome Res 2008;7:113–122. [PubMed: 18067247]
15. Liu H, Sadygov RG, Yates JR. Anal Chem 2004;76:4193–4201. [PubMed: 15253663]
16. Zybailov B, Mosley AL, Sardi ME, Coleman MK, Florens L, Washburn MP. J Proteome Res 2006;5:2339–2347. [PubMed: 16944946]
17. Mdluli K, Slayden RA, Zhu Y, Ramaswamy S, Pan X, Mead D, Crane DD, Musser JM, Barry CE. Science 1998;280:1607–1610. [PubMed: 9616124]
18. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaiia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Nature 1998;393:537–544. [PubMed: 9634230]
19. Barry C 3rd, Lee R, Mdluli K, Sampson A, Schroeder B, Slayden R, Ying Y. Prog Lipid Res 1998;37:143–179. [PubMed: 9829124]
20. Smith DM, Kafri G, Cheng Y, Ng D, Walz T, Goldberg AL. Mol Cell 2005;20:687–698. [PubMed: 16337593]
21. Lamichhane G, Raghunand T, Morrison N, Woolwine S, Tyagi S, Kandavelou K, Bishai W. J Infect Dis 2006;194:1233–1240. [PubMed: 17041849]
22. Darwin HK, Ehrt S, Gutierrez-Ramos JC, Weich N, Nathan CF. Science 2003;302:1963–1966. [PubMed: 14671303]
23. Martínez-Ruiz A, Villanueva L, González de Orduña C, López-Ferrer D, Higuera MA, Tarín C, Rodríguez-Crespo I, Vázquez J, Lamas S. Proc Natl Acad Sci U S A 2005;102:8525–8530. [PubMed: 15937123]
24. Rhee KY, Erdjument-Bromage H, Tempst P, Nathan CF. Proc Natl Acad Sci U S A 2005;102:467–472. [PubMed: 15626759]
25. Kumar A, Toledo JC, Patel RP, Lancaster JR, Steyn AJ. Proc Natl Acad Sci U S A 2007;104:11568–11573. [PubMed: 17609369]
26. O'Toole R, Smeulders MJ, Blokpoel MC, Kay EJ, Loughheed K, Williams HD. J Bacteriol 2003;185:1543–1554. [PubMed: 12591871]

27. Darwin HK. Nat Rev Microbiol 2009;7:485–491. [PubMed: 19483713]
28. Shi S, Ehrt S. Infect Immun 2006;74:56–63. [PubMed: 16368957]
29. Bryk R, Lima CD, Erdjument-Bromage H, Tempst P, Nathan C. Science 2002;295:1073–1077. [PubMed: 11799204]

Fig. 1.

A simplified comparison of the eukaryotic ubiquitination (A) and prokaryotic PUPylation (B) proteasomal degradation pathways. In eukaryotes, the backbone C-terminal carboxylate of ubiquitin is ligated to the target lysine *via* E1, E2 and E3 ubiquitin ligases followed by subsequent proteasomal degradation. In prokaryotes, the side chain amine of the C-terminal glutamine is deamidated by Dop followed by ligation to the target lysine by PafA. The AAA-ATPase facilitates translocation of PUPylated protein into the proteasomal core where the protein is degraded.

Fig. 2. Purification of PUPylated proteins from cell lysates. SDS-PAGE of His₇ tagged and non-tagged proteins purified by Ni-NTA affinity chromatography using both denaturing and native conditions. Samples were visualized by staining with Coomassie blue.

Fig. 3. Spectral annotations showing the characteristic losses of the deamidated glutamine and Q*GG motif from the target lysine. Of the 52 proteins exhibiting the +243 Da mass shift that could be manually annotated, 76% displayed MS2 fragment ions corresponding to these two distinctive losses.

Fig. 4.

The percent distribution of peptides for each protein within the His₇-tagged (under both native and denaturing conditions) and the non-tagged control samples. The confirmed PUPylated proteins (green circles) show decreased abundance as the number of peptides present in the non-tagged sample increases indicating a possible correlation between peptide distribution in the His₇-PUP and non-tagged control samples with the probability that a protein is PUPylated. Since no protein containing more than 44.6% of its peptides in the non-tagged control was confirmed to be PUPylated, any protein above this limit (as indicated by the black and yellow dashed line) was not considered as a possible target (red triangles) while any protein below this limit was considered to be a candidate protein (blue circles). Of the 193 proteins lacking a +234 Da modified peptide, 103 were found to be below the limit (candidate proteins), 78 were found to be above the limit (non-PUPylated proteins) and 12 were found to be mainly in

the His₇-PUP sample purified under native conditions suggesting they form protein complexes with PUPylated proteins (yellow triangles).

Fig. 5.

An example of a heavily regulated gene cluster in the mycolic acid biosynthetic pathway. Confirmed PUPylated proteins are shown in blue, candidate PUPylated proteins in pink and proteins not found in our study in grey.

Table 1

A list of the 52 confirmed protein targets for PUPylation. All spectra were initially screened using a 1% false positive discovery threshold as well as a *p*-value of not more than 0.05. Mass spectra exhibiting the characteristic +243 Da mass shift were confirmed through manual annotation for each protein (see Fig. SF1–SF48, ESI[†]). Protein targets only identified from InSpecT searches on clustered data sets are denoted by *

Protein name	GI Number	%Coverage	# Peptides	Total spectra	PUPylation modification site
[Mn]superoxide dismutase	118 469 767	93.2	79	937	HHATYVK+243GVNDALAK HHATYVK+243GVNDA HHATYVK+243GVN NLSPNGGDK+243PTGELAAAIIDDQFGSFDK
2,5-Diketo-D-gluconic acid reductase A*	118 173 010	28.7	3	5	IASNFDVDFELSGQDITSIASLETGK+243R
30S Ribosomal protein S17	118 171 465	56.1	5	8	LVEILEK+243AK*
30S Ribosomal protein S6	118 171 449	66.7	6	34	DGGTVDK+243VDIWGR
50S Ribosomal protein L11	118 473 904	26.8	4	24	VAK+243VTWDQVR
Acetolactate synthase 3 regulatory subunit	118 470 606	54.7	6	47	GFNIQSLAVGATEQK+243DMSR
Acetyl-CoA acetyltransferase*	118 170 768	72.3	19	58	AAAANK+243DGVFADEVVPVSIQR
Acetylornithine aminotransferase	118 472 927	35.6	11	103	AGVLGK+243TSLSHGIEELGHPLVVK
Aconitate hydratase 1	118 471 676	55.6	50	474	K+243DIHNYVEQNHPTPETK
Acyl carrier protein	118 471 527	34.3	5	61	TVGVVA YIQK+243LEENPEAAAAALR
Acyl-CoA synthetase	118 470 672	42.6	25	168	NVVVLGPGSIPK+243TPSGK
Adenylosuccinate synthetase	166 232 470	47.6	24	163	VSGGPPELTFDEHGAYLAK+243TGGEVGVTTGR
Alcohol dehydrogenase, zinc-containing	118 468 910	46.7	15	116	AMGADVTVLSQSLK+243K
Alkylhydroperoxide reductase*	118 172 773	64.6	18	59	VTSK+243DYEKG
Anthraniolate synthase component I*	118 472 702	40.3	14	27	GATEEEDVLEK+243ELLADEKER
Aspartate transaminase	118 46 8930	48.8	27	178	LGESK+243IASWTDPK
ATPase, AAA family protein	118 470 801	43.7	18	110	TLVTGK+243SSSASR
Branched-chain amino acid aminotransferase	118 472 496	31.5	13	46	KIDVDEWQK+243K
Carbonic anhydrase	118 169 435	58.7	7	39	*PNSNPVAAWK+243ALK
Carveol dehydrogenase*	118 169 959	25.1	5	6	ISSNEDIPASTPEDLAETVELVK+243GLNR
Chaperonin GroEL	118 469 510	57.1	28	361	VTETLLK+243SAK
Conserved hypothetical protein*	118 172 453	52.9	5	7	AIEWYGADV K+243
Conserved hypothetical protein	118 171 476	78.3	15	88	KDLIEAVEAEEPATEAETAQIK+243 DLIEAVEAEEPATEAETAQIK+243

Protein name	GI Number	%Coverage	# Peptides	Total spectra	PUPylation modification site
Cyclopropane-fatty-acylphospholipid synthase	118 173 720	67.5	33	447	EQATWAQK+243AIAQEGLTDLAEVR
DNA topoisomerase IV subunit B	118 468 752	33.8	15	41	SVTIWAANPANADTVTLWSK+ 243TGGGEVGVTTGR
Elongation factor Ts	166 221 229	67.6	20	178	NALVEADGDFDK+243A VELLR EADGDFDK+243AVELLR
Ferritin family protein	118 468 959	65.2	14	168	TNSGALDTK+243FHAIQDQIR
Glucose-6-phosphate-1-dehydrogenase	118 171 568	23.4	8	15	IIGVAK+234AGLDDAGYR
Glutamine synthetase, type I	118 474 021	31.8	12	61	GVEK+243GYVLGPQAEEDNVVSLTQEER
Glutamine synthetase, type I	118 169 243	51.7	24	189	LIK+243DENVEYVDIR
Glyceraldehyde-3-phosphate dehydrogenase, type I	118 467 799	54.4	15	81	VPIPTGSVTDLTAELAK+243SASVEDINAAMK
Glyoxalase family protein*	118 173 889	65.5	4	4	TADVDAAAK+243EIEATGGAIIK
Hypothetical protein	118 468 878	37.7	9	17	ADVIDEAAATFLK+243QQQEDK
Hypothetical protein	118 473 705	65.7	2	6	TLEENQK+243VEFEVQSPK
Inositol-5-monophosphate dehydrogenase	118 473 142	24.8	5	12	IAQVVEVAAK+243EPEPSAAIR
Malonyl CoA-acyl carrier protein transacylase	118 471 928	23.8	4	19	NAAGQIVAAAGAV AALDK+ 243LAEDPPAK
Myo-inositol-1-phosphate synthase	118 470 644	68.3	33	543	DVNFVAAFDDVDAK+243K
N-Carbamoyl-L-amino acid amidohydrolase	118 473 294	34.5	11	39	ELEK+243NNVTIGLVDR
Phosphoenolpyruvate carboxykinase*	118 173 553	65.8	30	100	LGYNVGDYFAHWINVVGK+243NADESK FLGYNVGDYFAHWINVVGK+243NADESK
Prolyl-RNA synthetase	118 472 270	56.6	29	147	DSYSFDVDDGLK+243NAYYQHR
Putative adenosylhomocysteinase	116 266 924	60.2	55	538	LSK+243DQAEYIGVDVEGYPY LSK+243DQAEYIGVDVEGYPKPEHYR
Putative thiosulfate sulfurtransferase	118 472 788	60.3	18	152	DFVDAQQFSK+243LLSER
Pyruvate dehydrogenase subunit E1	118 470 577	66.0	75	454	TK+243ALVADMSDQEIWNLK TK+243ALVADMSDQEIWNLKR
Ribonucleoside diphosphate reductase, alpha subunit*	118 170 953	41.2	23	58	K+243NEDMYLFSPPYDVER
Rieske 2Fe-2S family protein	118 474 008	33.4	12	35	DHNETVVLDFPK+243R
S-Adenosylmethionine synthetase	118 169 294	77.2	61	656	AAPVGLFVETFGSETVDPK+243IEK PVGLFVETFGSETVDPK+243IEK
Serine hydroxymethyltransferase	118 468 818	69.8	40	386	QPAFQQYAAQVADNAQALADGFVK+ 243R AAQVADNAQALADGFVK+243R
Short chain dehydrogenase	118 473 731	86.5	39	818	RDDLEVAAK+ 243ELDIESIVFDNTDASLEAVR
SseC protein	118 468 699	98.9	10	229	ALSSAGNGNVTVAPTGAGIHEVDVK+243VA*

Protein name	GI Number	%Coverage	# Peptides	Total spectra	PUylation modification site
Threonine synthase	118 468 890	66.7	18	148	K+243LTADFFPTIALVNSVNPYR
Transaldolase	118 171 487	57.4	18	80	LAHDTDKTILQAIELWK+243IVDRPNLLIK
Universal stress protein family protein	118 471 745	39.5	6	20	TLIDASK+243SAQMVVVGNR