Mycobacteria use a unique system for covalently modifying proteins based on the conjugation of a small protein, referred to as prokaryotic ubiquitin-like protein (PUP). In this study, we report a proteome-wide analysis of endogenous pupylation targets in the model organism Mycobacterium smegmatis. On affinity capture, a total of 243 candidate pupylation targets were identified by two complementary proteomics approaches. For 41 of these protein targets, direct evidence for a total of 48 lysine-mediated pupylation acceptor sites was obtained by collision-induced dissociation spectra. For the majority of these pupylation targets (38 of 41), orthologous genes are found in the M. tuberculosis genome. Interestingly, approximately half of these proteins are involved in intermediary metabolism and respiration pathways. A considerable fraction of the remaining targets are involved in lipid metabolism, information pathways, and virulence, detoxification and adaptation. Approximately one-third of the genes encoding these targets are located in seven gene clusters, indicating functional linkages of mycobacterial pupylation targets. A comparison of the pupylome under different cell culture conditions indicates that substrate targeting for pupylation is rather dynamic.
To explore the impact of pupylation as a new type of functional regulation in mycobacterial biology, it is essential to explore to what extent components of the mycobacterial proteome are targeted for pupylation. In this study, we used the endogenous PUP conjugation/processing machinery from *M. smegmatis* to identify potential PUP targets from enriched purified material. Using two complementary proteomics approaches, we have observed a total of 48 pupylation sites in 41 mycobacterial substrates. The majority of these targets are encoded by gene loci clustered in restricted regions of the *M. smegmatis* genome.

**Results and discussion**

**Identification of PUP substrates from the *M. smegmatis* proteome**

To provide insight into the overall properties of PUP, we first expressed the *M. tuberculosis* protein (Rv2111c, *mt*PUP) in *Escherichia coli* and purified it to homogeneity (Figure 1A). Further biophysical characterization is described in the Supplementary information.

To identify potential PUP substrates, we also expressed N-terminally poly-histidine-tagged *mt*PUP in *M. smegmatis*, which is a well-established model system (Hatfull et al., 2008). Affinity-purified *mt*PUP gave rise to multiple bands when assessed by SDS–PAGE (Figure 1B, Supplementary Figure 1A and B) and by western blot analysis (Supplementary Figure 1C). When using a *mt*PUP(Q64A) variant the multiple band pattern was lacking, indicating that it is caused by specific pupylation involving Gln64 of PUP. The same result was obtained when using a *M. tuberculosis* control target (Rv3874). Supporting this observation, *mt*PUP heterologously expressed in *E. coli* also migrated as a single band after nickel-nitrilotriacetic acid (NiNTA) affinity purification, showing that the effect is specific to *M. smegmatis* (Figure 1A).

We first used in-solution digestion followed by two-dimensional (2D) liquid chromatography, both by offline strong cation exchange (SCX) followed by online reversed phase (RP), and by electrospray ionization (ESI) tandem mass spectrometry (MS/MS), using a quadrupole-time-of-flight instrument (QqTOF). The aim of this approach was to obtain a high coverage of potential pupylation targets. NiNTA eluate protein fractions from *M. smegmatis* served as source material for this analysis. To minimize co-enrichment of indirect binders, the purification was carried out under denaturing conditions. To permit the recognition of unspecific binders to the NiNTA matrix, a parallel purification was carried out side-by-side from *M. smegmatis* cells containing the plasmid without an inserted *mt*PUP gene. To rule out the possibility that differences in the lists of observed proteins are the result of run-to-run variance during the SCX and RP separations, we incorporated isobaric tagging for relative and absolute quantification (iTRAQ) of the control and specific eluate digests into the workflow. The 2D liquid chromatography served to reduce sample complexity. Only two proteins (GroEL2, MSMEG_1583; CFP29, MSMEG_5830) were found in negative control experiments as well, probably because of the presence of sequence segments with an ability to bind to the NiNTA matrix with high affinity. The corresponding peptide peaks were subtracted from the data set.

Cumulatively, the analysis led to the identification of 1661 peptides that were observed only in the specific sample. They could be assigned to 243 putative pupylation targets, referred to as ‘candidate targets’ (Supplementary Table 1). For a protein to be considered as a target, it had to be identified based on the confident assignment of at least two strong collision-induced dissociation (CID) spectra. The analysis allowed a direct assignment of pupylation sites to 32 mycobacterial target proteins, referred to as ‘validated targets’ (Table I). For one of these, acyl carrier protein (MSMEG_4326), two distinct pupylation sites were observed. All pupylated peptides that were identified contained an internal lysine residue, which, due to the loss of the positive charge as a result of the bulky PUP conjugation to the side chain e-amine, lost the ability to function as a tryptic cleavage site (Figure 2A and B). This observation independently corroborated the evidence for correct pupylation assignment.

In parallel, the same samples were subjected to a complementary 2D gel electrophoresis analysis (Figure 2C). Single spots from the 2D gel were trypsinized and the resulting peptides were analyzed by ESI MS/MS (Supplementary Table 2). This approach led to the identification of 133 peptides that could be assigned to 13 validated pupylation substrates, covering a total of 17 confirmed pupylation sites (Table I). Three substrates were observed to be pupylated on more than one single lysine site: elongation factor Ts (MSMEG_2520, three sites, total number of lysines: 24), acyl carrier protein (MSMEG_4326, two sites, total number of lysines: 5) and alkylhydroperoxide reductase (MSMEG_4391, three sites, total number of lysines: 11).

For further analysis, we pooled the identified pupylated protein sequences and sorted them with respect to the gene loci organization in *M. smegmatis* (see Table I for pooled validated targets and Supplementary Table 3A for pooled candidate targets). The resulting list comprises 41 validated mycobacterial pupylation targets with a total of 48 non-redundant lysine-mediated pupylation sites. Owing to the direct demonstration of this type of lysine conjugation, these
| Target no | 2D Gel spots | Gene locus (M. smegmatis) | Detected pupylation sequences | Pupylated lysines | Gene locus (M. tuberculosis) | Pupylation site conserved | Functional annotation | Functional categorization | Gene cluster |
|-----------|--------------|--------------------------|------------------------------|------------------|----------------------------|--------------------------|------------------------|--------------------------|-------------|
| 1         | 1633         | MSMEG_0024               | HTYFGVEVDEESQK             | K147             | MSMEG_0009                | Poppyl-prolyl cis-trans isomerase | 2                      | Extracellular soluble-binding protein, family protein | 5            |
| 2         | 632          | MSMEG_0643               | YADTPYIPSYQDESYTQYQR       | K124             | MSMEG_0643                | Y 60-kDa chaperonin 1       | 0                      | Elongation factor Tu | 2            |
| 3         | 278          | MSMEG_0140               | WWK<sup>2</sup>VEELAMEWASIPDPR | K188             | MSMEG_0140                | Y Adenylation kinase      | 7                      | Y Adenylation kinase | 7            |
| 4         | 195          | MSMEG_1484               | LSGK<sup>2</sup>LLGIPQISTGDLFR | K23              | MSMEG_1484                | Y 30S ribosomal protein S11 | 2                      | 7            |
| 5         | 1685         | MSMEG_1522               | GVSASA<sup>2</sup>NKVEELAINDGOWACAR | K280             | MSMEG_1522                | Y Electron transfer flavoprotein, alpha subunit | 7                      | 7            |
| 6         | 1553         | MSMEG_1523               | VADVLRGRK<sup>2</sup>DAADVVR | K288             | MSMEG_1523                | Y 3-hydroxypropylmalate dehydrogenase | 7                      | 7            |
| 7         | 1712         | MSMEG_2387               | VVADVLRGRK<sup>2</sup>DAADVVR | K288             | MSMEG_2387                | Y Isopropylmalate isomerase small subunit | 7                      | 7            |
| 8         | 17          | MSMEG_2388               | NDSPFLNLGSPDK<sup>2</sup>GGVLAPDGIQOSSR | K51              | MSMEG_2388                | Y Isopropylmalate isomerase small subunit | 7                      | 7            |
| 9         | 80 (29.3)    | MSMEG_2520               | NAIVATTPQDK<sup>2</sup>AYELR | K86              | NAIVATTPQDK<sup>2</sup>AYELR | Y Elongation factor FtsI    | 2                      | Y Elongation factor FtsI | 2            |
| 10        | 1601        | MSMEG_2937               | ATTFYDDPDVLAK              | K275             | MSMEG_2937                | Y Pyridoxal biosynthesis lyase | 7                      | Y Pyridoxal biosynthesis lyase | 7            |
| 11        | 510         | MSMEG_2938               | LSYLQK                     | K172             | MSMEG_2938                | Y 3-oxoacyl-(acyl-carrier-protein) synthase 1 | 1                      | 1            |
| 12        | 1500        | MSMEG_3205               | GGVDVDAVVPK                | K188             | MSMEG_3205                | Y 3-oxoacyl-(acyl-carrier-protein) synthase 2 | 1                      | 1            |
| 13        | 1244        | MSMEG_3461               | FAPINSWDDVLAK              | K188             | MSMEG_3461                | Y Catalase-peroxidase 2     | 0                      | 0            |
| 14        | 47 (10.6)    | MSMEG_3526               | TVGTDVAAVKQ<sup>2</sup>KEEGFPAEAALR | K90              | MSMEG_3526                | Y Universal stress protein family protein | 10                     | 10           |
| 15        | 11          | MSMEG_3601               | QPAIEGFYK                  | K58              | MSMEG_3601                | Y Steroid delta isomerase | 3                      | 3            |
| 16        | 1898        | MSMEG_4326               | ETTLEK                     | K218             | MSMEG_4326                | Y Acetyl-CoA acetyltransferase | 1                      | 1            |
| 17        | 15          | MSMEG_4527               | RAQVLVPELEK                | K158             | MSMEG_4527                | Y Formate dehydrogenase     | 7                      | 7            |
| 18        | 1866        | MSMEG_4530               | IAPNK<sup>2</sup>PEE             | K218             | MSMEG_4530                | Y Sulfate ABC transporter, ATP-binding protein | 3                      | 3            |
| 19        | 1859        | MSMEG_4920               | AAA-AW<sup>2</sup>QVK           | K80              | MSMEG_4920                | Y Acetyl-CoA acetyltransferase | 1                      | 1            |
| 20        | 15          | MSMEG_5104               | NSLSGA<sup>2</sup>QEVVK     | K299             | MSMEG_5104                | Y Formate dehydrogenase     | 7                      | 7            |
| 21        | 1898        | MSMEG_5335               | WHPPDPAATVY<sup>2</sup>QVQGASG<sup>2</sup>SERF | K53              | MSMEG_5335                | Y Serine hydroxymethyltransferase | 7                      | 7            |
| 22        | 28          | MSMEG_6008               | ETLEK<sup>2</sup>QVQGASG<sup>2</sup>MAAAL | K218             | MSMEG_6008                | Y Acetyl-CoA acetyltransferase | 1                      | 1            |
Identification of mycobacterial pupylation targets

C Poulsen et al

Table I

| Target no. | Gene locus (M. smegmatis) | Gene locus (M. tuberculosis) | D2O-gel experiment | App. MW | Gene function | Functional category | Pupylation of site(s) | Detected pupylation sequence | Pupylated lysines | Pupylation sites | Peptide index | Peptide index (M. tuberculosis) |
|------------|-------------------------|-----------------------------|---------------------|--------|--------------|--------------------|-----------------------|-------------------------------|-----------------|----------------|----------------|--------------------------------|
| 17         | MSMEG_4298              | Rv2750c                      | V                   | 10     | ELSADFDLDAFEHV | AK                  | K242                  | Putative uncharacterized protein | PupylGGEWNSDEERIEMWLR | Rv3701c | Putative uncharacterized protein | 1                              |
| 18         | MSMEG_6247              | Rv3720c                      | Y                   | 38     | EQATWAQK | PupylGGEAIAQEGLTDLAEVR | K247                  | Y Cyclopropane-fatty-acyl-phospholipid synthase | 10               |
| 19         | MSMEG_6284              | Rv3722c                      | Y                   | 40     | LGESK | PupylGGEIASWTDPK | K339                  | Y Aspartate transaminase | 10               |
| 37         | MSMEG_6427              | Rv3846c                      | 0                   | 40     | HHATYVK | PupylGGEGVNDAIAK | K38                  | Y [Mn] superoxide dismutase | 0               |

Protein targets with pupylated sequences that have been identified by both methods: ESI QqTOF analysis and 2D gel analysis, are highlighted in gray. Column annotation: Global MS, peptide index from ESI-QqTOF analysis (Supplementary Table 1); 2D gel, peptide index from 2D gel electrophoresis (Supplementary Table 2); Gene locus (M. smegmatis), putative uncharacterized protein; Gene locus (M. tuberculosis), putative uncharacterized protein.

Table I (Continued)

| Target no. | Gene locus (M. smegmatis) | Gene locus (M. tuberculosis) | D2O-gel experiment | App. MW | Gene function | Functional category | Pupylation of site(s) | Detected pupylation sequence | Pupylated lysines | Pupylation sites | Peptide index | Peptide index (M. tuberculosis) |
|------------|-------------------------|-----------------------------|---------------------|--------|--------------|--------------------|-----------------------|-------------------------------|-----------------|----------------|----------------|--------------------------------|
| 38         | MSMEG_6427              | Rv3700c                      | V                   | 39     | DSADFLDAFEHV | AK                  | K232                  | Putative uncharacterized protein | PupylGGEWNSDEERIEMWLR | Rv3701c | Putative uncharacterized protein | 10               |
| 39         | MSMEG_6427              | Rv3700c                      | V                   | 40     | DSADFLDAFEHV | AK                  | K232                  | Putative uncharacterized protein | PupylGGEWNSDEERIEMWLR | Rv3701c | Putative uncharacterized protein | 10               |
| 40         | MSMEG_6427              | Rv3700c                      | V                   | 41     | DSADFLDAFEHV | AK                  | K232                  | Putative uncharacterized protein | PupylGGEWNSDEERIEMWLR | Rv3701c | Putative uncharacterized protein | 10               |

Genetic and functional clustering of pupylation targets

Although the genomes of several M. smegmatis strains have been sequenced (http://www.ncbi.nlm.nih.gov/nuccore/CP000480), no systematic functional annotation and categorization has yet been published. However, the vast majority (36/41) of our validated M. smegmatis targets with pupylation sites is mirrored by homologous proteins in M. tuberculosis H37Rv (Table I). Therefore, we have made use of available analyses of the related genome from the M. tuberculosis H37Rv strain, in which the proteome was divided into ten different functional categories (Camus et al., 2002). A total of 19 of these targets (47%) with homologs in M. tuberculosis H37Rv are involved in intermediary metabolism and respiration pathways. Substantial numbers of targets are also found for categories ‘lipid metabolism’ (six targets), ‘virulence, detoxification, adaptation’ (four targets) and ‘information path-
ways' (five targets). These categories are thought to be rich in potential targets that could become useful for future drug discoveries (Zhang, 2005).

When expanding this analysis to the complete list of *M. smegmatis* candidate targets we have identified as candidate pupylation substrates (Supplementary Table 3B), we notice that many of our observations for either overrepresentation or underrepresentation of specific functional target categories in the *M. tuberculosis* proteome serving as potential pupylation substrates are preserved (Figure 3). In our analysis, there is a distinct accumulation of potential pupylation substrate targets involved virulence, detoxification and adaptation (category 0), lipid metabolism (category 1), information pathways (category 2) and intermediary metabolism and regulation (category 7), when
compared with the overall presentation of targets from these categories within the entire proteome. In contrast, the number of potential pupylation substrate targets involved in cell wall and cell processes (category 3) is small, compared with its overall presentation within the complete proteome. Finally, we have observed no evidence for any pupylation substrates that are identified as targets in functional categories 4, 5 and 6.

We have also carried out an analysis of the pupylated targets in terms of gene locus organization. Strikingly, a substantial fraction of the identified targets with identified conjugation sites (13/41) is located in six gene clusters, I–VI (Table I), containing at least two genes encoded within one operon or in close proximity. This observation suggests that protein target pupylation could be correlated with functional linkages (Dandekar et al., 1998; Overbeek et al., 1999; Pellegrini et al., 1999). Noticeably, gene locus clustering of potential pupylation targets becomes even more obvious when the complete list of candidate pupylation targets is taken into account (Supplementary Table 3A).

Outlook

The availability of the complete genome and the corresponding proteome of M. tuberculosis has revolutionized fundamental research and generated new approaches to study disease mechanisms (Cole et al., 1998; Mattow et al., 2001). However, little is known about posttranslational modifications of targets from the M. tuberculosis proteome and related mycobacteria, except for phosphorylation (Greenstein et al., 2005; Wehenkel et al., 2008). In this contribution, we have shown for M. smegmatis that a substantial number of proteins may serve as targets for covalent modification by PUP conjugation, ultimately leading to alteration of their functional status and perhaps their fate for controlled proteasome-mediated degradation.

An important task for the future is to experimentally map the available evidence for pupylation of M. smegmatis targets onto the proteome of pathogenic M. tuberculosis strains and other pathogenic mycobacteria. Furthermore, it will be of specific interest to determine to what extent pupylation of targets may vary during different stages of the mycobacterial life cycle and within the host environment; this will require studies in appropriate in vivo models. Future studies exploring the significance of pupylation for fundamental biological processes in mycobacteria will be needed to elucidate aspects of pupylation biology that may find applications in ongoing efforts to overcome tuberculosis and other mycobacterial diseases.

Materials and methods

M. smegmatis growth under different experimental conditions

Bacterial growth was monitored by measuring the optical densities at 600 nm (OD\textsuperscript{600}) as a function of time at a wavelength of 600 nm, in triplicates. Cultures were prepared in identical triplicates for each time point, thereby ensuring that the growth of bacteria was not disturbed until the measurement time. Each cell density with an OD\textsuperscript{600} value exceeding 1.5 was diluted to allow density measurements within the linear range of the detector. Specific stress conditions were generated, by adding 2 mM H\textsubscript{2}O\textsubscript{2}, 3 mM NaNO\textsubscript{2} (pH 5.5) or 20 \textmu M epoxomicin (Enzo Life Sciences) to the growth medium.

Expression and purification of mtPUP

The ORF of PrcB, PrcA and PUP (Rv2109c, Rv2110c and Rv2111c) was amplified from the H37Rv genomic DNA by PCR, using the forward primer 5′-GAGCCATGCGCCGACAGCACCAA-3′ and the reverse primer 5′-GAGACGCTCTACGAGCGCTGCC-3′. For cloning of the mtPUP(Q64A) variant, the forward primer 5′-GAGGACTTTCGAGCGCAGTGGCC-3′ and reverse primer 5′-GCGGGATCCTTTATCAAGCTCCG-3′ were used. The PCR-introduced restriction sites,
EIS QqTOF mass spectrometry analysis for identification of PUP-conjugated targets

Protein-containing fractions were denatured in the presence of 6 M urea, 20 mM NaH₂HCO₃ (pH 8.0), followed by reduction with 1 mM Tris-(2-carboxyethyl)-phosphine for 30 min at 60°C and alkylation with 2.5 mM 4-vinylpyridine for 1 h at room temperature in the dark. Samples were diluted four-fold to ensure that the concentration of urea did not exceed 1.5 M. Tryptic digestion was initiated by the addition of 1% (wt/wt) of side chain-modified, TPKC-treated porcine trypsin and allowed to proceed at 37°C for 6 h.

Individual iTRAQ labeling reagents (Applied Biosystems, Foster City, CA, USA) were reconstituted in ethanol, added to peptide mixtures derived from the tryptic digestion of NiNTA eluate fractions (control, iTRAQ 114; sample, iTRAQ 115) and incubated at room temperature in the dark for 3 h.

The column effluent was coupled directly via a fused silica capillary transfer line to a QSTAR XL hybrid quadrupole/time-of-flight tandem mass spectrometer (Applied Biosystems; MDS Sciex). Raw iTRAQ ratios were corrected for impurity levels of individual reagent lots, determined by the manufacturer. Alternatively, peak lists for database searching were created using Mascot Distiller (Version 1; MatrixScience, London, UK), and searches were performed using designated MS/MS data interpretation algorithms within Mascot (Version 2.2; MatrixScience). Modifications considered were the attachment of GGG or GGE motifs to lysine residues, assuming a condensation reaction that proceeds with the concomitant loss of a water molecule and possible deamidation. In a negative control data analysis, a similar modification was assumed. For ProteinPilot searches, the algorithm assumes by default all possible unexpected cleavages and many of the more common modifications. Searches further considered up to two missed cleavages and charge states ranging from +2 to +4. In the few instances in which confidence values for assigned pupylation sites were not meeting or exceeding 95% confidence levels, the lower confidence, and thus higher assumed error rate, was partially offset by the presence of corroborating missed cleavages at the pupylated lysine residues and manual inspection of CID spectra. In particular, the latter approach led in a subset of spectra to the detection of candidate pupylation fragments that increased the confidence of assignments but were not considered by the algorithms for the calculation of scores. The experimental data have been submitted to the PRIDE data base (accession number: 11999).

For the analysis of 2D gels, spots, these were manually isolated from the gel, washed three times with 10 mM ammonium bicarbonate buffer (pH 7.8) and in-gel digested with trypsin (Promega, Mannheim, Germany) at 37°C overnight. For nano-HPLC/EIS-QqTOF analysis, tryptic peptides were extracted twice from gel pieces with 50% ACN in 0.1% trifluoroacetic acid. The MS/MS analyses were obtained on a high-capacity ion trap system (HCT ultra, Bruker Daltonics), in conjunction with an online RP nano-HPLC system (Dionex U3000, Dionex LC Packings, Idstein, Germany). The mass spectrometer was operated in the sensitive mode with the following parameters: capillary voltage, 1400 V; end plate offset, 500 V; dry gas, 8.01/min; dry temperature, 160°C; ion charge control, 15000; maximal fill-time, 500 ms. The nano-ESI source (Bruker Daltonics) was equipped with distal coated silica tips (FS360–20–10-D; New Objective). The MS spectra were recorded from the sum of seven individual CID spectra. In particular, the latter approach led in a subset of spectra to the detection of candidate pupylation fragments that increased the confidence of assignments but were not considered by the algorithms for the calculation of scores. The experimental data have been submitted to the PRIDE data base (accession number: 11999).

For the analysis of 2D gels, spots, these were manually isolated from the gel, washed three times with 10 mM ammonium bicarbonate buffer (pH 7.8) and in-gel digested with trypsin (Promega, Mannheim, Germany) at 37°C overnight. For nano-HPLC/EIS-QqTOF analyses, tryptic peptides were extracted twice from gel pieces with 50% ACN in 0.1% trifluoroacetic acid. The MS/MS analyses were performed on a high-capacity ion trap system (HCT ultra, Bruker Daltonics), in conjunction with an online RP nano-HPLC system (Dionex U3000, Dionex LC Packings, Idstein, Germany). The mass spectrometer was operated in the sensitive mode with the following parameters: capillary voltage, 1400 V; end plate offset, 500 V; dry gas, 8.01/min; dry temperature, 160°C; ion charge control, 15000; maximal fill-time, 500 ms. The nano-ESI source (Bruker Daltonics) was equipped with distal coated silica tips (FS360–20–10-D; New Objective). The MS spectra were recorded from the sum of seven individual CID spectra. In particular, the latter approach led in a subset of spectra to the detection of candidate pupylation fragments that increased the confidence of assignments but were not considered by the algorithms for the calculation of scores. The experimental data have been submitted to the PRIDE data base (accession number: 11999).
For protein identification, raw MS/MS data were searched using the Mascot algorithm (Matrix Science, v.2.2.0) against the NCBI nr database. The searches were performed with a mass tolerance of ± 1 Da for parent ions and ± 0.4 Da for fragment ions, considering for database searches. The reliability of protein identification was verified manually; a Mascot score > 60 was required to consider protein identification as significant. The experimental data have been submitted to the PRIDE data base (accession number: 12078).

2D gel electrophoresis

For 2D gel electrophoresis, protein eluate from the NiNTA preparation was treated three times with a buffer (30 mM Tris, 7.0 M urea, 2 M thiourea and 4% CHAPS (pH 8.5)) compatible with isoelectric focusing using Microcons (Millipore, Schwalbach, Germany) with a 3-5% cut-off. For IEF, a 20-μg sample was prepared by adding 0.7 μL DTT (1.08 g/ml; Bio-Rad) and 0.7 μL Servalyt 2–4 (Serva). Carrier ampholyte-based IEF was performed in a self-made IEF chamber, using tube gels (20 cm × 1.5 mm) as described elsewhere (Sitek et al., 2006). Briefly, after running a 21-h voltage gradient, the ejected tube gels were incubated in equilibration buffer (1.25 mM Tris, 40% (w/v) glycerol, 3% (w/v) SDS and 65 mM DTT (pH 6.8)) for 10 min. The second dimension was performed in a Desaphor VA 300 system (Sarstedt, Nürnberg, Germany) using polyacrylamide gels (15.2% total acrylamide and 1.3% bisacrylamide) as described elsewhere (Sitek et al., 2006). Therefore, the IEF tube gels were placed onto the polyacrylamide gels (20 cm × 30 cm × 1.5 mm) and fixed using 1.0% (wt/vol) agaro containing 0.01% (wt/vol) bromophenol blue dye (Riedel de Haen, Seelze, Germany). Silver staining was performed using an MS-compatible protocol (Blum et al., 1986).

To analyze the differential expression of proteins under different growth conditions, we have used the 2D DIGE method (Unlu et al., 2007). The secondary structure prediction was carried out using the computational server of Fred and PROOF (Quiall and King, 2000).

Multiple PUP sequence alignment

All PUP homologs of mtPUP (Rv211) were obtained in a global BLAST search. All unique sequences below an E-value of 1.0 were selected for the alignment. The alignment was carried out with ClustalX (2.0.9; Larkin et al., 2007). The secondary structure prediction was carried out using the computational server of Fred and PROOF (Quiall and King, 2000).

Functional annotation and gene clustering

The identifiers of M. smegmatis and M. tuberculosis genes have been taken from the Comprehensive Microbial Resources (http://cmr.jcvi.org). The identification of genes with paralogs in M. smegmatis and M. tuberculosis were carried out with the Artemis Comparison Tool (Carver et al., 2005). Functional annotation and categorization (M. tuberculosis) has been retrieved from Tuberculist (http://tuberculist.epfl.ch). Category identifiers were defined by (Camus et al., 2002). Pupylation targets were grouped in one gene cluster according to the following criteria: (a) genetic vicinity, for example, consecutive location and/or encoded in the same operon; and (b) conservation of gene order in M. smegmatis and M. tuberculosis.

Supplementary information

Supplementary information is available at the Molecular Systems Biology website (www.nature.com/msb).

Acknowledgements

YA thanks DAAD (German Academic Exchange Service) for a long-term doctoral fellowship. We thank Morlin Milewski for technical assistance. This study was supported by grants from the Canadian Institutes of Health Research (CIHR; MOP-74734 (GSU)), from BMBF (Pathogenomics-Plus, PTJ-BIO 031380L1) and from the European Commission Framework VII (NATT, 222965) to MW.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Blum H, Beier H, Gross HJ (1986) Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. Electrophoresis 8: 93–99

Burns KE, Liu WT, Bosshoff HI, Dorrestein PC, Barry III CE (2009) Proteasomal protein degradation in Mycobacteria is dependent upon a prokaryotic ubiquitin-like protein. J Biol Chem 284: 3069–3075

Camus JC, Pryor MJ, Medigue C, Cole ST (2002) Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv. Microbiology 148: 2967–2973

Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J (2005) ACT: the artemis comparison tool. Bioinformatics 21: 3422–3423

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry III CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S et al (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393: 537–544

Crosas B, Hanna J, Kirkpatrick DS, Zhang DP, Tone Y, Hathaway NA, Buecker C, Leggett DS, Schmidt M, King RW, Gygi SP, Finley D (2006) Ubiquitin chains are remodeled at the proteasome by opposing ubiquitin ligase and deubiquitinating activities. Cell 127: 1401–1413

Dandekar T, Snel B, Huynen M, Bork P (1998) Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci 23: 324–328

Darwin KH, Ehrt S, Gutierrez-Ramos JC, Weich N, Nathan CF (2003) The proteasome of Mycobacterium tuberculosis is required for resistance to nitric oxide. Science 302: 1963–1966

Daigelat S, Kowall J, Mattow J, Bumann D, Winter R, Hurwitz R, Kaufmann SH (2003) The RDI proteins of Mycobacterium tuberculosis: expression in Mycobacterium smegmatis and biochemical characterization. Microbes Infect 5: 1082–1095

Dummler A, Lawrence AM, de Marco A (2005) Simplified screening for function studies of Ser/Thr and Tyr protein phosphorylation in E. coli. Trends Biochem Sci 30: 1491–1494

Grabbe C, Ditkic I (2009) Functional roles of ubiquitin-like domain (ULD) and ubiquitin-binding domain (UBD) containing proteins. Chem Rev 109: 1481–1495

Greenstein AE, Grunder C, Echols N, Gay LM, Lombana TN, Mieczkowskia CA, Pullen KE, Sung PY, Alber T (2005) Structure/function studies of Ser/Thr and Tyr protein phosphorylation in Mycobacterium tuberculosis. J Mol Microbiol Biotechnol 9: 167–181

Hatfull GF, Cresawn SG, Hendrix RW (2008) Comparative genomics of the mycobacteriophages: insights into bacteriophage evolution. Res Microbiol 159: 332–339
Hershko A, Ciechanover A (1986) The ubiquitin pathway for the degradation of intracellular proteins. *Prog Nucleic Acid Res Mol Biol* 33: 19–56, 301

Hershko A, Ciechanover A (1998) The ubiquitin system. *Annu Rev Biochem* 67: 425–479

Iyer LM, Burroughs AM, Aravind L (2008) Unraveling the biochemistry and provenance of pupylation: a prokaryotic analog of ubiquitination. *Biol Direct* 3: 45

Kremer L, Dover LC, Carrere S, Nampoothiri KM, Lesjean S, Brown AK, Brennan PJ, Minnikin DE, Locht C, Besra GS (2002) Mycolic acid biosynthesis and enzymic characterization of the beta-ketoacyl-ACP synthase A-condensing enzyme from *Mycobacterium tuberculosis*. *Biochem J* 364: 423–430

Kusunose E, Ichihara K, Noda Y, Kusunose M (1976) Superoxide dismutase from *Mycobacterium tuberculosis*. *J Biochem* 80: 1343–1352

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948

Ma Q, Zhao X, Edde AN, Geerlof A, Li X, Cronan JE, Kaufmann SH, Wilmanns M (2006) The *Mycobacterium tuberculosis* LipB enzyme functions as a cysteine/lysine dyad acyltransferase. *Proc Natl Acad Sci USA* 103: 8662–8667

Mattow J, Jungblut PR, Schaible UE, Mollenkopf HJ, Lamer S, Zimny-Arndt U, Hagens K, Muller EC, Kaufmann SH (2001) Identification of proteins from *Mycobacterium tuberculosis* missing in attenuated *Mycobacterium bovis* BCG strains. *Electrophoresis* 22: 2936–2946

Ouali M, King RD (2000) Cascaded multiple classifiers for secondary structure prediction. *Protein Sci* 9: 1162–1176

Overbeek R, Fonstein M, D’Souza M, Pusch GD, Maltsev N (1999) The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci USA* 96: 2896–2901

Pearce MJ, Mintseris J, Ferreyra J, Gygi SP, Darwin KH (2008) Ubiquitin-like protein involved in the proteasome pathway of *Mycobacterium tuberculosis*. *Science* 322: 1104–1107

Pellegrini M, Marcotte EM, Yeates TO (1999) A fast algorithm for genome-wide analysis of proteins with repeated sequences. *Proteins* 35: 440–446

Rappasiber J, Ichihamura Y, Mann M (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* 75: 663–670

Sitek B, Potthoff S, Schulenberg T, Stiegbauer J, Vinke T, Rump LC, Meyer HE, Vonend O, Stuhler K (2006) Novel approaches to analyse glomerular proteins from smallest scale murine and human samples using DIGE saturation labelling. *Proteomics* 6: 4337–4345

Stiegitz KA, Yang H, Roberts MF, Stee B (2005) Reaching for mechanistic consensus across life kingdoms: structure and insights into catalysis of the myo-inositol-1-phosphate synthase (mIPS) from *Archaeoglobus fulgidus*. *Biochemistry* 44: 213–224

Stiegbel F, Imkamp F, Sutter M, Steiner M, Mamedov A, Weber-Ban E (2009) Bacterial ubiquitin-like modifier Pup is deamidated and conjugated to substrates by distinct but homologous enzymes. *Nat Struct Mol Biol* 16: 647–651

Unlu M, Morgan ME, Minden JS (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18: 2071–2077

Wahlenke A, Bellinzozi M, Grana M, Duran R, Villarino A, Fernandez P, Andre-Leroux G, England P, Takiff H, Cervenansky C, Cole ST, Aizari PM (2008) Mycobacterial Ser/Thr protein kinases and phosphatases: physiological roles and therapeutic potential. *Biochim Biophys Acta* 1784: 193–202

Zhang Y (2005) The magic bullets and tuberculosis drug targets. *Annu Rev Pharmacol Toxicol* 45: 529–564

Molecular Systems Biology is an open-access journal published by European Molecular Biology Organization and Nature Publishing Group. This work is licensed under the Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.