Int J Clin Exp Med (2009) 2, 309-328
www.ijcem.com/IJCEM910001

Original Article

Acid stress response of a mycobacterial proteome: insight from a gene ontology analysis

Bryan A.P. Roxas[1,3], Qingbo Li[1,2]

[1]Center for Pharmaceutical Biotechnology, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60607, USA; [2]Department of Microbiology and Immunology, College of Medicine, University of Illinois at Chicago, Chicago, IL 60612, USA; [3]Current address: Department of Veterinary Science and Microbiology, University of Arizona, Tucson, AZ 85721, USA.

Received October 7, 2009; accepted November 5, 2009; available online November 10, 2009

Abstract: Acidity in vesicles of macrophages is a general signal that bacteria respond to during infection. Mycobacteria are particularly capable of resisting the acidification in macrophages that engulf the bacteria. In this work, we used label-free quantitative proteomics to study the *Mycobacterium smegmatis* proteome under acid stress so as to gain an insight into the acidic adaptation in mycobacteria. We quantified 1032 proteins. With a 3-fold change threshold, 20 and 52 proteins were found regulated at false discovery rates of 5% and 14% respectively. We performed a systems analysis based on gene ontology for the global proteome expression profile. We found that the most significant changes induced by the acid stress include a downregulation of transmembrane transporter activity and an upregulation of enzymes involved in fatty acid metabolism. The results suggest that reduced transmembrane transport and increased fatty acid metabolism probably contribute to or associate with acid tolerance in mycobacteria.

Key words: Mycobacterium smegmatis, acid stress, label-free proteomics, gene ontology, systems biology, membrane transport

Introduction

During bacterial infection, macrophages become activated to result in a series of events specifically designed to induce killing of engulfed microorganisms. These events include the gradual acidification of phagosomes, phagolysosome fusion, induction of reactive oxygen and nitrogen intermediates, and antigen processing [1]. Mycobacterium species, especially *Mycobacterium tuberculosis* (Mtb), has evolved effective biochemical mechanisms to resist the killing by macrophages. Upon infection, mycobacteria sense and respond to the hostile intraphagosomal environment. The response leads to the synthesis and secretion of critical proteins or lipids to interfere with the host defense. Acidity is a dominant signal that mycobacteria respond to during invasion because the membrane-bound acidic vesicles are the primary weaponry for macrophages to eliminate invading pathogens. On the other hand, there is evidence that vesicles containing live Mtb are not acidic [2, 3], which suggests that Mtb possesses biochemical mechanisms to evade the bactericidal action of the phagolysosome. This ability for Mtb to resist the phagolysosome fusion, however, is not necessarily present during the initial entry into host cells. To the contrary, a majority of bacillus-containing phagosomes fuse with lysosomes within the first a few hours of infection. It was suggested that the fusion of Mtb-containing phagosomes with lysosomes was actually advantageous for the Mtb that infected macrophages [4]. Consistent with the phagolysosomal adaptation, Mtb survived at an in vitro pH of 4.5 in a simple buffer and maintained "intrabacterial pH" [5]. It was proposed that an escape from the phagolysosome played a role in Mtb-phagolysosome interaction [6]. Survival and proliferation was enhanced for bacilli emerging from fused phagolysosomes, and intracellular passage

increased the ability of Mtb to avoid phagolysosome fusion in a subsequent infection [7]. Taken together, the evidence suggests that an acid adaptation plays an important role in the interaction of mycobacteria with macrophages.

Mycobacterium smegmatis (Msm), a widely used model mycobacterium system to study mycobacterial biology and to screen anti-tuberculosis drugs [8-11], is non-pathogenic and is known to be eventually cleared from macrophages. It is interesting to note, however, that Msm also possesses the ability to initially halt the acidification of the phagosome within five hours after entry into the murine J774 macrophage [12]. After five hours, the intraphagosomal pH slowly acidifies. With live Msm uptaken by the macrophage, only 20–25% of the phagosomes acidified in the first hour post-infection. This level of phagosome acidification rose only slightly over the next seven hours. Between eight and 24 hours, most phagosomes became acidic to lead to an eventual killing of the bacteria by the 48th hour. In contrast, 80% of the phagosomes containing heat-killed Msm acidified rapidly after one hour. The acidification kinetics with the dead Msm was similar to that seen with latex-bead-containing phagosomes. These results indicate that live Msm initially withstands the acid assault and actively delays the acidification of phagosomes.

In this study, we used an advanced proteomics platform and bioinformatics methods to investigate Msm culture cells under acid stress to gain insight into the acidic adaptation of mycobacteria at the proteome level.

Materials and methods

Cell cultures

Msm strain mc^2 155 was obtained from the American Type Culture Collection (ATCC; Rockville, Md). Cells were cultured as described previously [13, 14]. Briefly, we grew a pH 5.0 and a pH 7.0 Msm culture in triplicate to mid-log phase and harvested at OD 0.7. The cultures were grown in 100-ml 7H9 medium under shaking at 37°C in loosely capped 250-ml nephelo culture flasks which had a 19-mm diameter side arm. An OD value was recorded at 600 nm in a Spectronic 20D spectrophotometer (Thermo Fisher Scientific, Waltham, MA). An aliquot of 30 ml was collected from each culture replicate and pelleted at 4000 rpm in a 5810R refrigerated Eppendorf centrifuge for 10 min at 4°C. A [^{15}N]-labeled culture was also grown to log-phase for use as an internal standard to determine false positive rates in protein quantitation [13]. Hereafter, we name the stressed pH 5 culture as **S**, the reference pH 7 culture as **R**, and the internal standard culture as **IS**.

Protein sample preparation

Each cell pellet was resuspended in 100 mM ammonium bicarbonate buffer and lysed by bead beating in the presence of a protease inhibitor cocktail (Pierce, Rockford, IL) [14]. The whole cell lysates were cleared by centrifugation at 13,000 g for 30 min at 4°C. Protein concentrations were quantified with the BCA protein quantitation kit (Pierce, Rockford, IL). The triplicate protein extracts for cultures S and R were pooled respectively. The pooled protein extracts from cultures S and R were mixed respectively with an equal amount of protein extract from the IS culture to generate two protein mixtures i.e., S_P (for culture S) and R_P (for culture R). One hundred micrograms of proteins from each of S_P and R_P were separated on a 10% Tris-HCl SDS-PAGE gel (Pierce) and fractionated into 5 fractions. Gel bands were processed for in-gel digestion and peptide extraction as described [13, 14].

LC/MS analysis of peptides

The peptide extracts from the gel bands were submitted for analysis with the nanoLC/LTQ-FTMS system (Thermo Finnigan; San Jose, CA) in the Proteomics and Informatics Services Facility (PISF) in the Research Resources Center at University of Illinois at Chicago supported by the Searle Funds at the Chicago Community Trust. Each peptide extract sample was analyzed with duplicate LC/MS injections as previously described [13].

Specifically, in each injection, 5 μL of peptide extract solution was separated on a 150-mm x 75-μm C18 reverse phase column with a 5% to 35% acetonitrile (v/v) gradient in 0.1% trifluoroacetic acid over 60 min. The LTQ-FTMS was operated in a data-dependent acquisition mode with up to 10 MS/MS spectra acquired following each MS scan. The acquired RAW data files were searched against the National Center for Biotechnology Information (NCBI) database of *M. smegmatis* strain mc^2 155 (downloaded in 2006 with the old locus

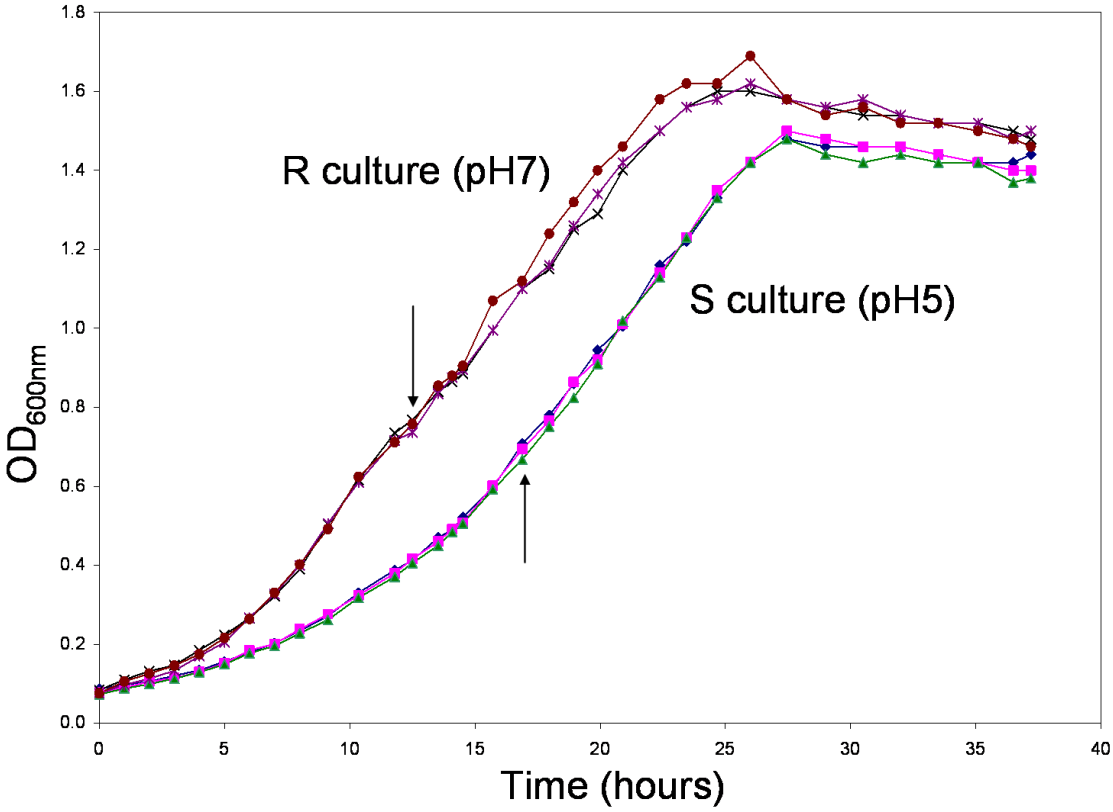

Figure 1. Growth curves of the S and R cultures in triplicate. The arrows indicate the timepoints to harvest the culture samples for proteomic analysis.

names) in two separate BioWorks searches. One search corresponded to [^{14}N] labeling and the other to [^{15}N] labeling. The precursor ion tolerance was set to ±1.5 Da. Trypsin was designated as the digestion enzyme with two missed cleavages allowed. Peptide and protein probabilities were calculated by BioWorks. Only peptides with P <.01 were accepted for subsequent quantitation of abundance (Table S1). After peptide and protein identifications, we carried out peptide quantitation using Matlab v7.2 (MathWorks, Natick, MA) and Microsoft Excel based on the previously described methods [13, 15, 16].

Protein quantitation

We quantified the protein abundances with a label-free proteomics approach as described previously [13, 15, 17]. The abundance of a protein was represented by the sum of the extracted ion chromatographic intensities of the peptide charge states detected for that protein. A peptide charge state is a peptide ionized to a specific charge in a mass spectrometer. One peptide could be detected at multiple peptide charge states [16]. We identified >5000 unique peptide charge states (p <.01). Only peptide charge states identified at p <.01 were used for subsequent quantitation of proteins. We only accept the proteins having peptide(s) identified by ≥2 peptide charge states identification events (p <.05).

From protein sample S_P, the abundance of a protein was quantified for its abundance in culture S (A_S) and its abundance in culture IS ($A_{IS,S}$). Likewise, from protein sample R_P, the abundance of a protein was quantified for its abundance in culture R (A_R) and in culture IS ($A_{IS,R}$). In total, each protein had four abundance values i.e., A_S, A_R, $A_{IS,S}$ and $A_{IS,R}$. We used A_S and A_R to determine the positives i.e., the differentially regulated proteins between cultures S and R. $A_{IS,S}$ and $A_{IS,R}$ represented replicate quantitation of the same proteins from culture IS. Thus, the differentially regulated proteins selected based on a difference

between $A_{IS,S}$ and $A_{IS,R}$ were false positives. The false discovery rate was the quotient of false positives over positives as applied previously [13].

Results

Cell culturing

The inoculants for the S and R cultures were pre-grown at pH 5 and pH 7 respectively to late log phase before use. About $2x10^9$ cells were used to inoculate each 100-ml medium to start the growth of the S and R cultures in triplicate (**Figure 1**). The cells were allowed to grow for approximately three doublings before being harvested at an OD value of about 0.7. This OD value corresponded to about OD 0.4/ml after adjustment for the side arm diameter of the nephelo culture flask. Three doublings are consistent with the typical number of doublings of a bacillus that infects a macrophage to result in a few bacilli in the macrophage [18].

The intraphagosomal pH decreases to <5.0 after uptake of dead mycobacteria or latex particles due to the fusion of phagosomes with lysosomes. The fused phagolysosome has been shown to have a pH ranging from 4.7 to 5.0 in mouse peritoneal macrophages and baby hamster kidney cells [19]. Typically, Mtb maintains the intraphagosomal pH between 6.2 and 6.6 [20]. When the mycobacterium-containing macrophages are activated by interferon-γ, the intraphagosomal pH could decrease to 5.2, at which point the activated macrophages become bacteriostatic or bactericidal. It was observed that most of the bacilli infecting macrophages were initially within fused vacuoles [6].

Thus, the pH of 5.0 that we chose in this study represents the low end of the pH that a mycobacterium could possibly encounter in a macrophage early in an infection [20]. In a previous study, we studied the immediate response of Msm to an acid shock with a protein turnover approach [21]. In this study, we examined the adaptive acid stress response of Msm in the cultures shown in **Figure 1**.

Protein quantitation

We quantified 1032 proteins (the list is available from the Author upon request). Each protein had four abundance values i.e., A_S, A_R, $A_{IS,S}$ and $A_{IS,R}$ (see Methods). **Figure 2** shows the overlay of two scatter plots based on these four protein abundance values for the 1032 proteins. One plots A_S versus A_R. The other plots $A_{IS,S}$ versus $A_{IS,R}$. The ratio of A_S versus A_R was used to select the proteins regulated in the acid stressed culture S versus the reference culture R. The ratio of $A_{IS,S}$ versus $A_{IS,R}$ was used to determine false positives due to the variations introduced during sample handling and data analyses.

Differentially regulated proteins

We determined the Msm proteins that were differentially expressed between the acid stressed culture S and the reference culture R. From the 1032 proteins, 70 unlabeled proteins were found regulated between the pH 5 and pH 7 growth conditions at a 3-fold change threshold. But 17 labeled internal standard proteins were also found "regulated" at the same threshold. The presence of the 17 false positives resulted in a false discovery rate of 24% which we considered too high for subsequent biological interpretation. If we used a 2-fold change threshold, the false discovery rate would be 40% even though the total number of differentially regulated proteins would increase to 162. Thus, we empirically chose the 3-fold change cutoff to select differentially regulated proteins.

As shown in the insert graph in **Figure 2**, there was an inverse relationship between the false discovery rate and the protein abundance level. Based on this relationship, we reduced the false discovery rate by applying an additional threshold of protein abundance level. Above an average protein abundance level of approximately $1x10^5$, the false discovery rate decreased to 14%. This 14% false discovery rate cutoff resulted in 52 differentialy regulated unlabeled proteins (**Table 1**). Further incorporation of additional lower-abundance proteins resulted in a significant increase of false discovery rate. Thus, we chose the 14% false discovery rate cut-off as a compromise between the acceptable error rate and the number of selected differentially regulated proteins. Sufficient numbers of differentially regulated proteins affords a statistical test in a system-based analysis. To interpret individual differentially regulated proteins, a ≤5% false discovery rate would be necessary. At a false discovery rate of 5%, 20 proteins were found regulated. These 20 high-confidence differen-

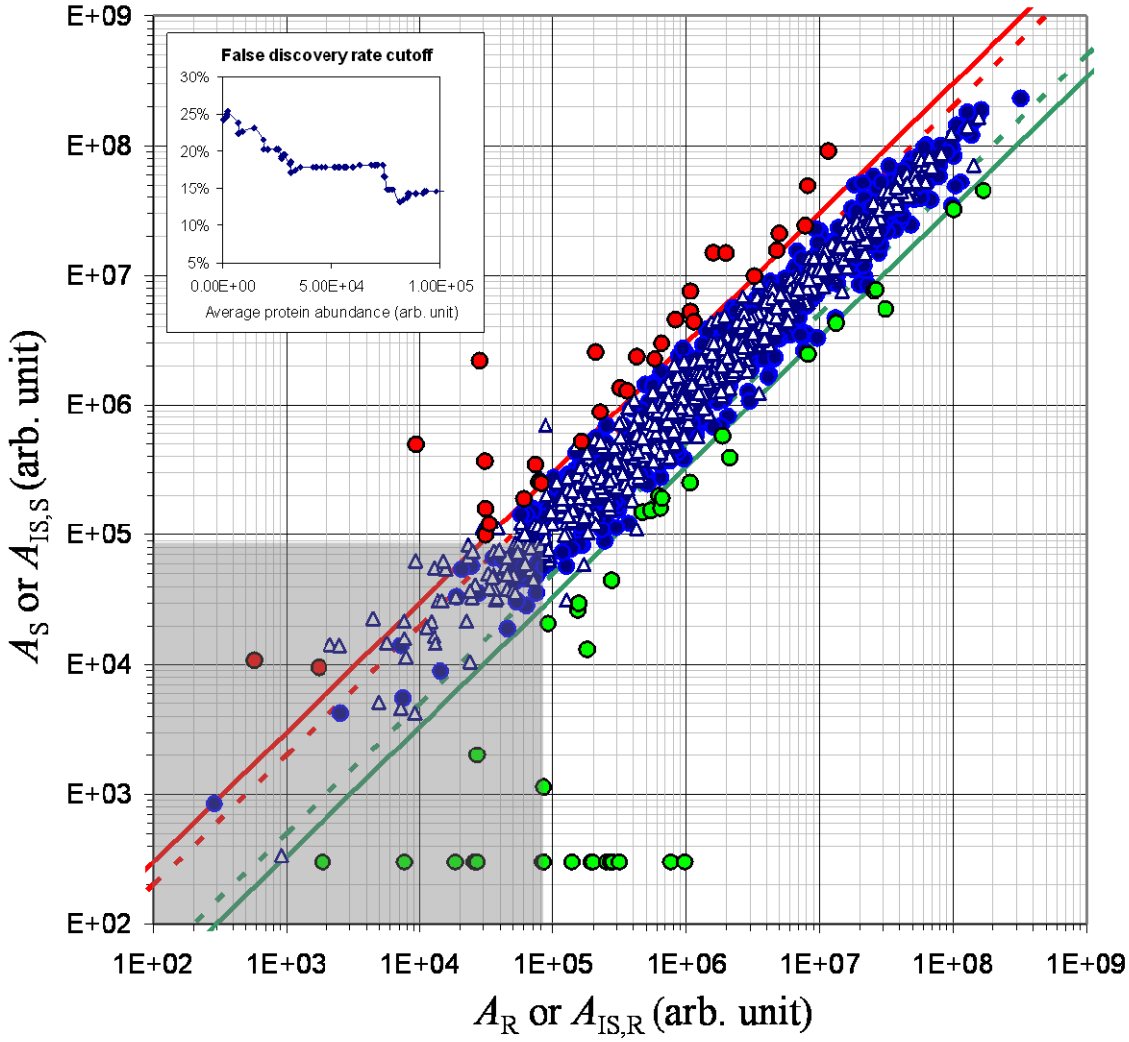

Figure 2. Selection of differentially expressed proteins from the 1032 detected proteins. Two scatter plots are overlaid. One corresponds to the A_S/A_R ratio (circular colored markers) and the other to the $A_{IS,S}/A_{IS,R}$ ratio (the triangular white markers). A red line indicates the upper threshold of 3-fold (solid) or 2-fold (dashed) change. A green line indicates the lower threshold of 3-fold (solid) or 2-fold (dashed) change. The red, blue, and green circular markers respectively represent the upregulated, unchanged, and downregulated proteins in the S (pH 5) *vs* the R (pH 7) culture. The insert at the upper left corner indicates the relationship between the false discovery rate and the average protein abundance in the low-abundance region as indicated with the grey shade.

tially regulated proteins are indicated in **Table 1**.

Thus, we identified 20, 52, and 70 differentially regulated proteins at estimated false discovery rates of 5%, 14%, and 24% respectively. For the systems analyses in the following, we used the 52 differentially regulated proteins selected at the 14% false discovery rate.

Interactions among the differentially regulated proteins

We first examined the interaction among the 52 differentially regulated proteins based on the evidence from the String database [22]. The purpose was to identify functionally related proteins that might be co-regulated under the acid stress condition. We identified two protein clusters with >2 proteins (**Figure 3**). One cluster was upregulated and the other

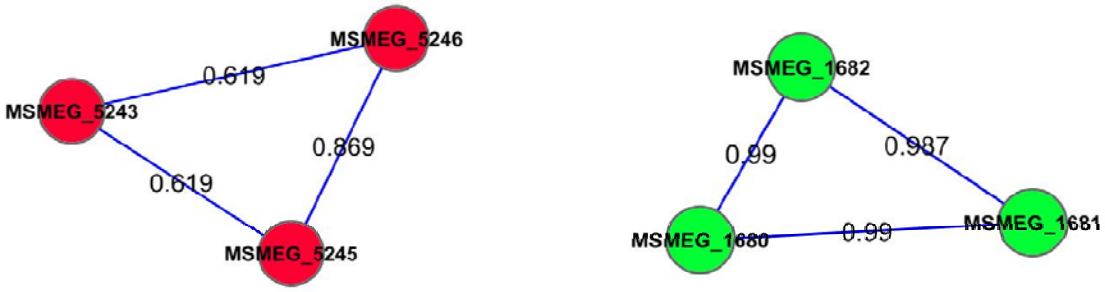

Figure 3. The two protein clusters identified among the 52 differentially regulated proteins. Red and green indicate up- and down-regulation respectively. The numbers on the edges indicate the interaction confidence scores.

downregulated. All of these six proteins were among the 20 high-confidence differentially regulated proteins selected at a false discovery rate of 5% (**Table 1**).

The three proteins in the upregulated protein cluster i.e., MSMEG_5243, MSMEG_5245, and MSMEG_5246 belong to a multi-gene locus encoding the DevR response regulator in Msm [23]. MSMEG_5245 is annotated as a universal stress protein (USP). It has a conserved USP domain and shares 50% similarity and 34% identity with Rv3134c that is the first gene in the Mtb *devR/devS* operon [24]. In Msm, downstream of MSMEG_5245 is MSMEG_5244 that shares 94% similarity and 85% identity with the Mtb *devR*. MSMEG_5244 is the only *devR* homologue in Msm. MSMEG_5244 was detected but not found to be regulated by the thresholds applied in this study. It was shown to be a stationary-phase regulator required for the adaptation of Msm to oxygen-starvation and resistance to heat stress [23].

MSMEG_5245 responded to an oxygen-starvation condition in a *devR*-dependant manner, but did not respond to a carbon-starvation condition or a UV stress [23]. An inhibition of the Msm aerobic respiration, however, did not induce the DevR regulon [25]. The inhibition of aerobic respiration under hypoxic conditions or nitric oxide induction did not result in an increase in the DevS kinase activity [11]. Although it was shown that the DevSR two-component system has a regulatory role during hypoxia-induced dormancy in Mtb, Msm, and BCG [26], the direct linkage of the DevSR two-component system to the functional state of the respiratory electron transport chain remains to be determined. In this study, MSMEG_5245 was upregulated in response to the acid stress along with another two UPS proteins i.e., MSMEG_3950 and MSMEG_3945 (**Table 1**). The result in this study suggests that the Msm DevSR two-component system also responds to an acid stress in addition to heat stress and oxygen starvation [23].

There are three proteins in the downregulated protein cluster including MSMEG_1680-1682. MSMEG_1680 is a conserved hypothetical protein. MSMEG_1681 is an endoribonuclease L-PSP superfamily protein that belongs to a widely distributed family of YER057c/YjgF/UK114 proteins of unknown function [27]. Members of the YER057c/YjgF/UK114 family of proteins are conserved among all domains of life. In bacteria, endoribonuclease L-PSP superfamily proteins were suggested to be potentially involved in the transformation or transport of small molecules, such as antibiotics [28, 29]. A rat endoribonuclease L-PSP superfamily protein was shown to inhibit protein translation at the initiation but not at the elongation stage [30]. The conservation of this protein in all domains of life suggests its fundamental function in a cell. MSMEG_1682 is a flavin-containing monooxygenase which is a complementary enzyme system to the cytochrome P450 family of enzymes and oxygenates several soft, highly polarizable nucleophilic heteroatom-containing chemicals and drugs [31].

Gene ontology analysis

To assess the major biological themes perturbed by the acid stress in Msm, we performed a gene ontology (GO) analysis for the 52 differentially regulated proteins (**Table 1**).

Table 1. The 52 differentially regulated proteins selected at 14% false discovery rate. The 20 proteins shown in bold belong to the high-confidence list of differentially regulated proteins selected at 5% false discovery rate.

Locus	Protein description	PCS_IDs	A_R	A_S	$A_{IS,R}$	$A_{IS,S}$	A_S/A_R
(*Upregulated proteins*)							
MSMEG_0614	Methyltransferase	2	3.58E+05	1.29E+06	1.79E+05	3.20E+05	3.61 [a]
MSMEG_0969	**Glutamate-1-semialdehyde-2,1-aminomutase (hemL)**	**8**	**3.22E+06**	**9.92E+06**	**1.72E+06**	**4.57E+06**	**3.08** [a]
MSMEG_1037	**Alcohol dehydrogenase, zinc-containing**	**61**	**1.16E+07**	**9.09E+07**	**5.86E+06**	**8.82E+06**	**7.82** [a]
MSMEG_1475	**Conserved hypothetical protein**	**4**	**6.54E+05**	**2.99E+06**	**6.56E+05**	**6.59E+05**	**4.57** [a]
MSMEG_1903	Caib-baif family	6	2.26E+05	8.85E+05	9.29E+04	6.94E+04	3.91 [a]
MSMEG_2669	**Hydrolase**	**46**	**5.03E+06**	**2.10E+07**	**4.72E+06**	**5.22E+06**	**4.17** [a]
MSMEG_2956	**NAD-dependent epimerase-dehydratase family protein**	**47**	**1.59E+06**	**1.49E+07**	**6.24E+06**	**7.72E+06**	**9.37** [a,b]
MSMEG_3932	14 kDa antigen	2	2.09E+05	2.56E+06	5.07E+04	4.75E+04	12.2
MSMEG_3945	**Universal stress protein family**	**14**	**1.08E+06**	**5.22E+06**	**4.01E+05**	**3.21E+05**	**4.83** [a]
MSMEG_3950	**Universal stress protein family**	**20**	**7.81E+06**	**2.41E+07**	**1.30E+06**	**1.43E+06**	**3.09** [a]
MSMEG_4381	Amidase	2	8.13E+04	2.50E+05	3.19E+05	5.63E+05	3.08 [a]
MSMEG_4971	Oxidoreductase	2	1.64E+05	5.19E+05	1.36E+06	9.37E+05	3.16 [a]
MSMEG_5136	**Helix-turn-helix motif**	**8**	**8.36E+05**	**4.53E+06**	**4.69E+04**	**6.80E+04**	**5.42** [a]
MSMEG_5164	Zinc-binding alcohol dehydrogenase family protein	6	3.17E+05	1.35E+06	3.48E+06	8.77E+06	4.27 [a]
MSMEG_5243	**Helix-turn-helix motif**	**4**	**1.14E+06**	**4.35E+06**	**2.33E+05**	**2.07E+05**	**3.82**
MSMEG_5245	**Universal stress protein family**	**8**	**1.08E+06**	**7.58E+06**	**3.06E+05**	**4.46E+05**	**7.02** [a]
MSMEG_5246	**Conserved hypothetical protein**	**22**	**1.98E+06**	**1.49E+07**	**2.94E+06**	**2.11E+06**	**7.53** [a]
MSMEG_5285	Phospholipase, patatin family	14	5.80E+05	2.26E+06	1.42E+05	2.21E+05	3.90 [a]
MSMEG_5419	**Putative lipoprotein**	**66**	**4.80E+06**	**1.56E+07**	**3.26E+06**	**3.81E+06**	**3.24**
MSMEG_5739	Putative long-chain fatty-acid--CoA ligase	16	4.23E+05	2.36E+06	3.81E+05	5.32E+05	5.57 [a,b]
MSMEG_6454	**Conserved hypothetical protein**	**56**	**8.18E+06**	**4.89E+07**	**3.58E+06**	**3.41E+06**	**5.98** [a]
(*Down-regulated proteins*)							
MSMEG_0131	AMP-binding enzyme, putative	2	9.20E+04	2.08E+04	9.30E+04	7.72E+04	0.23 [a]
MSMEG_0317	Conserved hypothetical protein	4	1.87E+06	5.74E+05	8.30E+05	4.95E+05	0.31
MSMEG_0394	Hypothetical protein	27	1.07E+06	2.53E+05	6.73E+05	4.81E+05	0.24
MSMEG_0987	Hypothetical protein	2	7.68E+05	3.00E+02	1.45E+06	1.49E+06	0.00
MSMEG_1052	Hypothetical protein	12	1.81E+05	1.30E+04	3.36E+05	3.90E+05	0.07 [a,b]
MSMEG_1445	30S ribosomal protein S17	2	1.94E+05	3.00E+02	4.34E+04	5.69E+04	0.00 [a]
MSMEG_1605	Phosphate transport system regulatory protein PhoU (phoU)	2	8.48E+04	1.13E+03	9.13E+03	4.25E+03	0.01

(Table 1 continued)

MSMEG_1680	**Conserved hypothetical protein**	**20**	**8.20E+06**	**2.46E+06**	**3.89E+06**	**4.36E+06**	**0.30**
MSMEG_1681	**Endoribonuclease L-PSP superfamily**	**156**	**2.57E+07**	**7.57E+06**	**1.90E+07**	**2.68E+07**	**0.29**
MSMEG_1682	**Flavin-containing monooxygenase FMO**	**62**	**2.66E+07**	**7.76E+06**	**1.12E+06**	**1.34E+06**	**0.29** [a,b]
MSMEG_1832	Conserved hypothetical protein	2	3.14E+05	3.00E+02	4.89E+05	4.74E+05	0.00
MSMEG_1951	Conserved domain protein	14	4.67E+05	1.50E+05	1.29E+06	1.41E+06	0.32
MSMEG_2116	PTS system, glucose-specific IIBC component	6	1.56E+05	2.98E+04	2.63E+05	1.72E+05	0.19 [a,b]
MSMEG_2789	Acetyltransferase, GNAT family	2	1.39E+05	3.00E+02	1.45E+05	1.56E+05	0.00 [a]
MSMEG_2942	Glyoxalase-bleomycin resistance protein-dioxygenase superfamily protein	2	2.70E+05	3.00E+02	1.84E+06	1.98E+06	0.00 [a,b]
MSMEG_3962	**Lactate 2-monooxygenase**	**64**	**3.15E+07**	**5.44E+06**	**2.00E+07**	**2.37E+07**	**0.17** [a,b]
MSMEG_4075	CoA-binding protein	2	1.54E+05	2.65E+04	5.62E+05	3.66E+05	0.17 [a]
MSMEG_4298	**3-Methyl-2-oxobutanoate hydroxymethyltransferase (panB)**	**53**	**1.69E+08**	**4.48E+07**	**2.11E+06**	**1.36E+06**	**0.27** [a]
MSMEG_4476	Hypothetical protein	2	2.80E+05	3.00E+02	8.88E+04	2.04E+05	0.00
MSMEG_4935	**ATP synthase F1, epsilon subunit (atpC)**	**41**	**1.33E+07**	**4.26E+06**	**4.53E+06**	**3.98E+06**	**0.32** [a,b]
MSMEG_4936	**ATP synthase F1, beta subunit (atpD)**	**190**	**1.01E+08**	**3.19E+07**	**4.31E+07**	**4.67E+07**	**0.31** [a,b]
MSMEG_5006	Phosphohistidine phosphatase	2	8.29E+04	3.00E+02	3.33E+04	3.91E+04	0.00
MSMEG_5022	Flavin-containing monooxygenase FMO	12	5.38E+05	1.55E+05	8.05E+05	6.37E+05	0.29 [a,b]
MSMEG_5694	Conserved hypothetical protein	2	2.52E+05	3.00E+02	1.82E+06	1.28E+06	0.00
MSMEG_5703	Molybdenum cofactor biosynthesis protein C (moaC)	2	9.80E+05	3.00E+02	7.42E+05	1.16E+06	0.00 [a]
MSMEG_5835	Fumarate reductase-succinate dehydrogenase flavoprotein	2	6.58E+05	1.92E+05	1.70E+05	5.88E+04	0.29 [a]
MSMEG_6075	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ispF)	2	8.50E+04	3.00E+02	1.74E+05	2.78E+05	0.00 [a]
MSMEG_6159	Conserved domain protein	10	6.17E+05	2.01E+05	4.21E+05	1.10E+05	0.33 [a]
MSMEG_6210	Conserved hypothetical protein	2	2.75E+05	4.46E+04	9.40E+05	1.13E+06	0.16
MSMEG_6457	Oxidoreductase molybdopterin binding domain, putative	2	6.40E+05	1.61E+05	4.07E+05	3.96E+05	0.25 [a]
MSMEG_6913	Putative transcriptional regulatory protein	2	1.99E+05	3.00E+02	1.19E+05	1.67E+05	0.00 [a]

[a] Annotated in GO. [b] Present in the unique enriched GO terms (Table 2). PCS_IDs - the number of times peptides from a protein were identified by MS/MS scan.

Table 2. Enriched unique GO terms for the 52 differentially regulated proteins.[a]

GO_ID	*p*	x	n	GO term description	Proteins in test set
45261	0.006	2	3	Proton-transporting ATP synthase complex, catalytic core F(1)	MSMEG_4935; MSMEG_4936
5886	0.008	3	10	Plasma membrane	MSMEG_4935; MSMEG_2116; MSMEG_4936
15672	0.006	3	9	Monovalent inorganic cation transport	MSMEG_4935; MSMEG_1052; MSMEG_4936
42777	0.006	2	3	Plasma membrane ATP synthesis coupled proton transport	MSMEG_4935; MSMEG_4936
16701	0.01	2	4	Oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	MSMEG_2942; MSMEG_3962
4043	0.01	2	4	L-Aminoadipate-semialdehyde dehydrogenase activity	MSMEG_2956; MSMEG_5739
4497	0.01	3	12	Monooxygenase activity	MSMEG_1682; MSMEG_5022; MSMEG_3962
22892	0.001	6	30	Substrate-specific transporter activity	MSMEG_4935; MSMEG_1052; MSMEG_2116; MSMEG_4936; MSMEG_2956; MSMEG_5739
46961	0.006	2	3	Hydrogen ion transporting ATPase activity, rotational mechanism	MSMEG_4935; MSMEG_4936
15294	0.01	2	4	Solute:cation symporter activity	MSMEG_1052; MSMEG_2116

[a] *p* – the *p*-value of a hypergeometric test performed with BiNGO in Cytoscape.
x – the number of differentially regulated proteins in the test set of a GO term.
n – the number of detected proteins in a GO term.

The GO project provides controlled vocabularies for the description of the biological process, molecular function, and cellular component in a cell [32]. For the 1032 quantified Msm proteins, 835 were annotated in the GO annotation for Msm [33]. The GO annotations for these 835 Msm proteins were extracted using the BiNGO plug-in in Cytoscape [34]. The extracted GO annotations for these 835 proteins were used as the custom annotation file to determine the enriched GO terms for the 52 differentially regulated proteins (**Table 1**). We searched for enriched GO terms in the GO hierarchy consisting of the three branches including *biological_process*, *molecular_function*, and *cellular_component*.

Selection of enriched GO terms

For the 52 differentially regulated proteins, 37 were annotated in the GO and formed a 132-node network when analyzed with Cytoscape (data not shown). Of the 132 nodes (GO terms), 75 were enriched with differentially regulated proteins based on a hypergeometric test ($p<.05$) [34].

A close examination of the 75 enriched GO terms indicates that there was significant interdependence and redundancy among them. Many enriched GO terms had the same set of differentially regulated proteins. Such redundancy was mostly because we had a relatively small number of differentially regulated proteins for the enrichment test.

We were interested in the enriched GO terms that had their unique sets of proteins, or so-called unique enriched GO terms. If multiple enriched GO terms shared the same set of proteins and were in parent-child relationship along one branch of the GO hierarchy tree [35], we chose the one at the end of the branch as the unique enriched GO terms, as illustrated in the simplified GO network shown in **Figure 4**. In **Figure 4**, we only retained the unique enriched GO terms and their parents with ≥2 proteins. The unique enriched GO terms are indicated with a thicker border in **Figure 4**.

We identified 10 unique enriched GO terms (**Table 2**). Other enriched GO terms were the consequence of the enrichment of one or more of these 10 child terms. For example, although the last seven GO terms were enriched in the *biological_process* branch, the GO term *plasma membrane ATP synthesis coupled proton transport* was the first one to have MSMEG_4935 and MSMEG_4936 (**Table 2**). Another three GO terms upstream of it contained the same two proteins. They did not contribute to new information. Thus, they were not included in **Table 2**. Further up along the *biological_process* branch was the GO term *monovalent inorganic cation transport* that had an additional protein MSMEG_1052. Because this GO term included an additional protein different from those in its child terms, this GO term was included in **Table 2**. Another nine GO terms in **Table 2** were selected based on the same principle.

In **Figure 4**, the GO term *L-aminoadipate-semialdehyde dehydrogenase activity* contains only upregulated proteins. Another two GO terms i.e., *substrate-specific transporter activity* and *transporter activity* contain both up- and downregulated proteins. All other enriched GO terms contain only downregulated proteins. *Substrate-specific transporter activity* contains two upregulated and four downregulated proteins (**Table 2**). The enrichment of *transporter activity* is the direct result of the enrichment of its child term *substrate-specific transporter activity*.

Cellular component

The GO hierarchy branch *cellular_component* contains nine differentially regulated proteins (**Figure 4**). The enriched GO term *proton-transporting ATP synthase complex, catalytic core F(1)* contains AtpC (MEMEG_4935) and AtpD (MSMEG_4936). AtpC and AtpD are the subunits of the F1 motor. The F1 motor produces ATP in the presence of a proton gradient. It also participates in the regulation of intrabacterial pH homeostasis. The enriched GO term *plasma membrane* contains MSMEG_2116 in addition to AtpC and AtpD (**Table 2**). MSMEG_2116 is a glucose-specific IIBC component of the phosphoenolpyruvate (PEP): carbohydrate phosphotransferase system (PTS). It catalyzes the phosphorylation of N-acetyl-D-glucosamine (GlcNAc) to N-acetyl-D-glucosamine-6-phosphate (GlcNAc-6-p). The PEP:PTS system is involved in the transport of a large number of carbohydrates, in chemotaxis towards these carbon sources, and in the regulation of a number of other metabolic pathways [36]. The PTS catalyzes the uptake of carbohydrates and their conversion into their respective phosphoesters during trans-

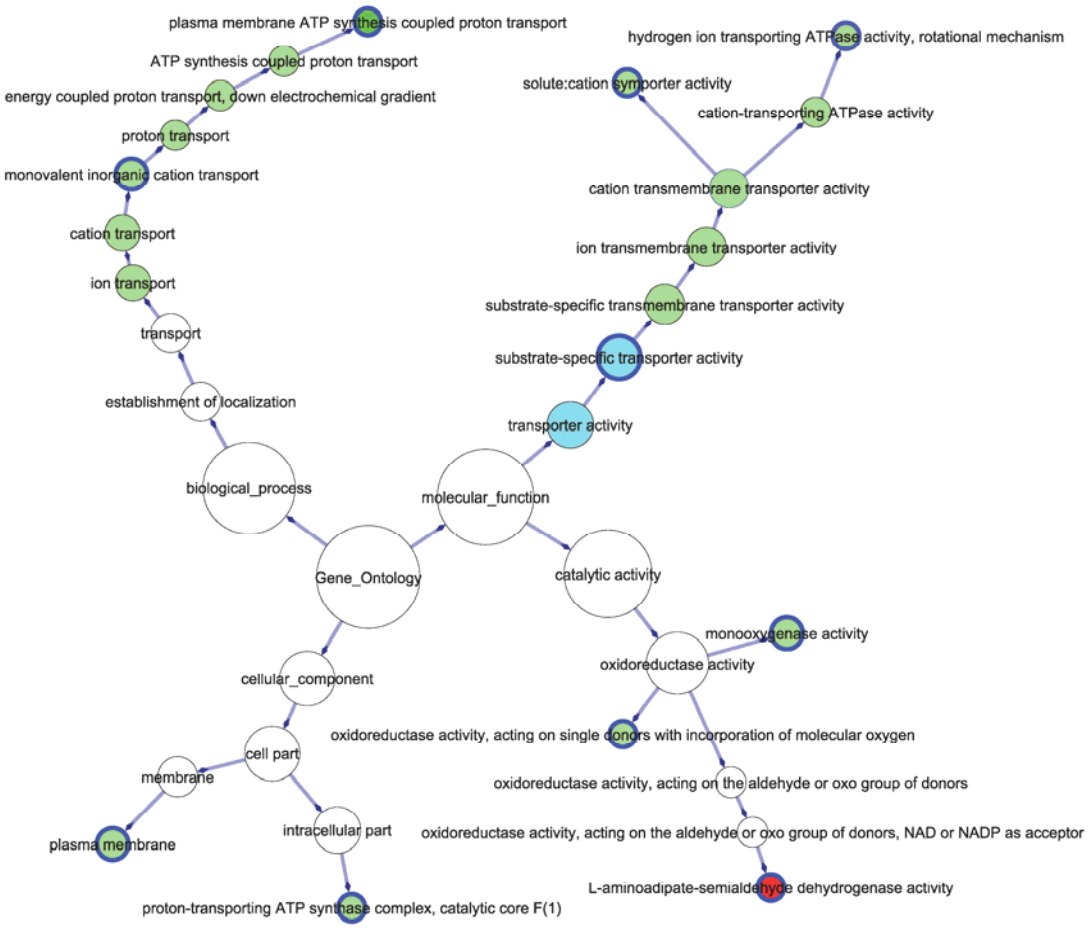

Figure 4. Enriched GO terms for the 52 differentially regulated proteins. The colored nodes are enriched GO terms. The nodes with a thicker border are unique enriched GO terms (see text). The size of a node is proportional to the number of proteins that it contains. A red node contains only upregulated proteins, and a green node contains only downregulated ones. A cyan node contains both up- and down-regulated proteins.

port. The result here suggests that the acidic growth condition affects the substrate uptake and energy production, consistent with the slower exponential phase growth of the S culture (**Figure 1**). It remains to be determined whether the reduced substrate uptake results in lower ATP synthesis or the lower ATP synthesis results in reduced energy-dependent substrate uptake. We also cannot exclude that these two processes are separately regulated by the acid stress and that one is not necessarily the consequence of the other.

Biological process

The GO hierarchy branch *biological_process* contains 29 differentially regulated proteins (**Figure 4**). The enriched GO term *plasma membrane ATP synthesis coupled proton transport* contains AtpC and AtpD. Again, the downregulation of this GO term suggests that ATP synthesis could be reduced under the acid stress. The GO term *monovalent inorganic cation transport* contains AtpC, AtpD, and MSMEG_1052. MSMEG_1052 is a hypothetical protein. The abundance of this hypothetical protein was downregulated by 14-fold (**Table 1**), suggesting that it has a role in the acid stress response. Indeed, the results in **Table 2** indicate that MSMEG_1052 was also annotated in several other enriched GO terms in the *molecular_function* ontology including *substrate-specific transporter activity*, *solute:cation symporter activity*, and *active*

transmembrane transporter activity. These results support a role of MSMEG_1052 in transporter activity. Interestingly, the String database [22] annotates MSMEG_1052 as an amino acid carrier protein and shows that it interacts with an alanine racemase (MSMEG_1575), a sodium/proline symporter (MSMEG_5303), a hexapeptide transferase family protein (MSMEG_1055), and an uncharacterized putative protein (MSMEG_1053). The protein interaction evidence suggests that MSMEG_1052 might be involved in the transport of amino acids. In this study, the down-regulation of MSMEG_1052 suggests a reduced amino acid transport in acid stressed Msm cells.

Molecular function

The GO hierarchy branch *molecular_function* contains 32 proteins (**Figure 4**). The enriched GO term *L-aminoadipate-semialdehyde dehydrogenase activity* contains two upregulated proteins i.e., MSMEG_5739 and MSMEG_2956. Annotated as an NAD-dependent epimerase-dehydratase family protein (**Table 2**), MSMEG_2956 also shares 55% identity and 72% similarity with MSMEG_5739 which is annotated as a putative long-chain fatty-acid-CoA ligase. Long-chain fatty-aid CoA ligases catalyze the bioactivation of fatty acids to form acyl-CoA thioesters. Acyl-CoA thioesters are used as substrates for subsequent metabolic pathways. The long-chain ligases exist as a super family of membrane proteins. They play key roles in both fatty acid activation and xenobiotic acyl-CoA formation [37].

The upregulation of MSMEG_5739 and MSMEG_2956 suggests that Msm increased the metabolism of fatty acids under the acid stress. A recent work by Deb et al. showed that Mtb accumulated triacylglycerol and wax ester under multiple stresses including acidity [38]. Thus, the upregulation of MSMEG_5739 and MSMEG_2956 in Msm under the acid stress implies that the cells likely increased anabolism of fatty acids. Accumulation of triacylglycerol and wax ester under stress could in turn result in a decrease in the cell wall/membrane permeability.

In the microarray study by Fisher et al. [39], the Mtb homologues of MSMEG_5739 and MSMEG_2956 were not upregulated by the acid shock in that experiment. Meanwhile, although the nonribosomal peptide synthetases/polyketide synthases were found most significantly induced in the acid shocked Mtb cells in the microarray study by Fisher et al. [39], the two detected proteins of their Msm homologues MSMEG_0408 and MSMEG_6392 were not regulated in this study. The duration of acid shock in the study by Fisher et al. was short. Thus, the acid shock condition in Fisher et al.'s study was somewhat different from the acid-adapted growth condition in this study. In addition, given the moderate correlation of microarray and proteomic data [40], a discrepancy between microarray and proteomic data should not be regarded purely as an experimental variation. This kind of discrepancy warrants a proteomic study to measure the abundance of a protein that is the combined result of transcription, translation, and other post-translational regulations such as degradation or turnover [41].

On the other hand, *fadD9* (Rv2590) and *nrp* (Rv0101), which are two Mtb homologues of MSMEG_2956, were upregulated in Mtb that infected activated murine bone marrow macrophages [42]. A dozen of other Mtb genes homologous to the upregulated Msm protein in **Table 1** were upregulated in the bacilli that infected activated macrophages, as will be discussed later.

Two downregulated proteins i.e., MSMEG_2942 and MSMEG_3962 are within the enriched GO term *oxidoreductase activity, acting on single donors with incorporation of molecular oxygen*. MSMEG_2942 is a glyoxalase/bleomycin resistance protein/dioxygenase superfamily protein. Glyoxalase I catalyzes the first step of the glyoxal pathway to convert methylglyoxal and reduced glutathione to *S*-lactoylglutathione. *S*-lactoylglutathione is then converted by glyoxalase II to lactic acid [43, 44]. The glyoxalase system is believed to have evolved to detoxify reactive 2-oxoaldehydes which include mainly methylglyoxal. Methylglyoxal is formed endogenously as a by-product of the triosephosphate isomerase reaction in glycolysis [45]. MSMEG_3962 is a lactate 2-monooxygenase that catalyzes the oxidation of sodium lactate by dioxygen to pyruvate. Pyruvate is then decarboxylated to acetate. The down-regulation of MSMEG_2942 and MSMEG_3962 could be due to the reduced transport of glucose into the cell under the acid stress, which in turn led to a lower glycolysis activity.

Table 3. Regulation of the homologous Mtb genes in intraphagosomal bacilli. These Mtb genes are homologues to the differentially regulated Msm proteins shown in Table 1. The mRNA relative abundances were determined for the bacilli that infected naïve and activated murine bone marrow macrophages for 4, 24, and 48 hours respectively [42].

Mtb gene	Mtb gene description	In naïve macrophages			In activated macrophages			Homologous Msm protein	Similarity (%)
		4 hr	24 hr	48 hr	4 hr	24 hr	48 hr		
(Mtb genes matched to the genes of the upregulated Msm proteins in Table 1)									
Rv3127	hypothetical protein	1.3	1.7	1.9	1.7	21.3	7.5	MSMEG_5246	60.00
acg	hypothetical protein	0.9	1.3	1.6	1.3	19.7	9.6	MSMEG_5246	65.24
acr	14 KD antigen	0.6	1.2	1.4	0.8	17.1	11.6	MSMEG_3932	78.32
Rv3129	hypothetical protein	1.1	0.6	1.3	1.3	11.8	7.8	MSMEG_5243	76.59
TB31.7	hypothetical protein	1.1	1.4	1.4	1.1	9.2	5.2	MSMEG_3945	60.49
Rv2005c	hypothetical protein	0.9	1.0	1.2	1.1	6.7	2.5	MSMEG_3950	65.52
Rv0893c	hypothetical protein	1.4	1.6	1.8	2.6	2.6	2.5	MSMEG_0614	71.19
Rv2026c	hypothetical protein	1.6	1.5	1.7	1.8	2.4	2.2	MSMEG_3945	60.99
Rv0725c	hypothetical protein	1.7	2.0	2.3	2.1	2.4	2.6	MSMEG_0614	61.36
Rv3767c	hypothetical protein	2.3	2.0	1.7	2.4	2.2	2.0	MSMEG_0614	60.60
fadD9	acyl-CoA synthetase	2.4	2.2	1.8	2.4	2.2	2.0	MSMEG_2956	82.08
nrp	peptide synthetase	1.3	1.6	1.6	1.5	2.1	1.7	MSMEG_2956	73.58
Rv1062	hypothetical protein	1.2	1.4	1.8	1.5	2.1	1.8	MSMEG_5285	72.93
Rv3787c	hypothetical protein	1.4	1.3	1.3	1.6	1.7	1.7	MSMEG_0614	62.79
Rv2781c	hypothetical protein	1.1	1.6	1.7	1.2	1.4	2.3	MSMEG_4971	60.42
Rv1889c	hypothetical protein	1.4	1.3	1.3	1.5	1.4	1.4	MSMEG_0614	62.39
Rv0281	hypothetical protein	0.7	1.2	1.7	0.7	1.3	1.7	MSMEG_0614	74.08
hemL	glutamate-1-semialdehyde 2,1-aminomutase	1.0	0.9	1.1	1.2	1.1	1.1	MSMEG_0969	85.42
Rv2765	hypothetical protein	1.0	0.8	0.9	1.0	0.9	0.9	MSMEG_2669	73.57
adhC	alcohol dehydrogenase	0.9	0.8	0.8	0.8	0.7	0.7	MSMEG_1037	74.06
(Mtb genes matched to the genes of the downregulated Msm proteins in Table 1)									
cspA	cold shock protein	0.5	0.6	0.4	0.4	0.3	0.3	MSMEG_6159	98.50
atpD	ATP synthase beta chain	0.6	0.4	0.4	0.5	0.3	0.3	MSMEG_4936	98.50
Rv0227c	hypothetical protein	0.6	0.5	0.4	0.5	0.3	0.3	MSMEG_0317	75.74

(Table 3 continued)

fadD5	probable fatty-acid CoA ligase	0.5	0.5	0.5	0.6	0.5	0.5	MSMEG_0131	84.58
panB	3-methyl-2-oxobutanoate hydroxymethyltransferase	1.0	1.2	1.0	0.9	0.7	0.7	MSMEG_4298	83.27
rpsQ	30s ribosomal protein s17	1.0	1.0	0.8	0.8	0.7	0.7	MSMEG_1445	90.72
moaC2	probable MoaC-2 protein involved in molybdopterin synthesis	0.9	1.1	1.1	0.8	0.7	0.8	MSMEG_5703	71.89
phoY1	probable phosphate transport system regulatory protein	0.9	0.8	0.7	0.9	0.8	0.7	MSMEG_1605	69.40
Rv0042c	hypothetical protein	0.9	0.9	0.8	0.9	0.9	0.7	MSMEG_6913	65.03
Rv3259	hypothetical protein	1.1	0.9	1.0	1.0	1.1	1.0	MSMEG_1832	92.08
phoY2	probable phosphate transport system regulatory protein	1.0	1.4	1.2	1.0	1.2	1.1	MSMEG_1605	76.05
Rv0875c	hypothetical protein	0.8	1.1	1.4	0.9	1.6	1.5	MSMEG_5694	74.05
ispF	hypothetical protein	1.4	1.6	2.1	1.6	1.7	2.3	MSMEG_6075	69.53
Rv0785	hypothetical protein	1.2	1.7	1.9	1.4	2.2	1.8	MSMEG_5835	82.95
Rv2669	hypothetical protein	1.3	1.5	1.6	1.4	2.3	1.7	MSMEG_2789	63.80

The enriched GO term *monooxygenase activity* contains three downregulated proteins including MSMEG_1682, MSMEG_5022, and MSMEG_3962. MSMEG_1682 and MSMEG_5022 share 30% identity and 48% similarity. Both are flavin-containing monooxygenases that metabolize many clinically important xenobiotic compounds as well as endogenous substrates as part of a discrete physiological process [46]. MSMEG_5022 shares 27% identity and 37% similarity with the Mtb monooxygenase EthA. The EthA transcript level determined by microarray analysis showed a down-regulation after four, 24, and 96 hours of starvation in a nutrient-starvation model of Mtb persistence [47]. EthA is responsible for the oxidative activation of the second-line anti-tubercular prodrugs ethionamide and thiacetazone [48].

Under the enriched GO term *substrate-specific transporter activity*, the unique enriched child term *hydrogen ion transporting ATPase activity, rotational mechanism* contains AtpC and AtpD. The unique enriched GO term *solute:cation symporter activity* contains MSMEG_1052 and MSMEG_2116. As described before, MSMEG_2116 is a glucose-specific IIBC component of the PEP:PTS system. The PEP:PTS system is involved in transport and phosphorylation of a large number of carbohydrates, in chemotaxis towards these carbon sources, and in regulation of a number of other metabolic pathways [36]. The result again suggests that the acidic growth condition affected the substrate uptake that requires energy. MSMEG_1052 is a probable amino acid carrier protein and is predicted to have a sodium:amino acid symporter activity based on its GO annotation [49]. A sodium:amino acid symporter catalyzes the transfer of an amino acid solute located outside a membrane, to enter with Na^+ into the cytoplasm. MSMEG_1052 contains the domain of the sodium:alanine symporter family [50]. The acidic growth condition appeared to repress the sodium:alanine transport, or amino acid transport in general.

Intraphagosomal regulation of Mtb genes homologous to the differentially regulated Msm proteins

To assess the state of the biochemical environment in phagosomes harboring Mtb, Schnappinger et al. captured the transcriptional responses of Mtb in macrophages before and after immunologic activation [42]. We examined whether the Mtb genes homologous to the Msm differentially regulated proteins in **Table 1** responded to the intraphagosomal environment. In **Table 3**, we compiled a list of the Mtb genes with homology to the Msm genes encoding the differentially regulated Msm proteins shown in **Table 1**. The matched genes between Mtb and Msm were selected with the multi-genome homology comparison tool from the J. Craig Venter Institute (www.jcvi.org). We arbitrarily set the gene similarity threshold at 60%. As shown in **Table 3**, 20 Mtb genes are homologous to 12 upregulated Msm proteins, and 15 Mtb genes are homologous to 14 downregulated Msm proteins. In **Table 3**, we incorporated the 3-timepoint mRNA relative expression ratios for the 35 Mtb genes in both naïve and activated murine wild-type bone marrow macrophages from Table S1 entitled "Regulation of All Analyzed Genes" in [42]. The Mtb genes were regarded differentially expressed when the mRNA relative expression ratio had a >2-fold change. The mRNA relative expression ratio was calculated as the mRNA abundance in intraphagosomal Mtb relative to that in Mtb from a log-phase 7H9 broth culture [41].

It is interesting to note that 14 out of the 20 Mtb genes homologous to the upregulated Msm proteins were upregulated in at least one time point in either or both of the naïve and activated macrophages. None of these 20 Mtb genes were downregulated at any timepoint in the infected macrophages. Meanwhile, 12 out of the 15 Mtb genes homologous to the downregulated Msm proteins were either downregulated (four) or unchanged (eight) in the macrophages (**Table 3**).

Discussion

In this study, we have utilized the advanced nanoLC/LTQ-FTMS proteomics system and bioinformatics methods to investigate the response of Msm to acid stress. Acid stress has an implication in the interaction of mycobacteria with macrophages. Msm is a fast-growing mycobacterium whose adaptive response to hypoxia and nitric oxide exposure is similar to that of Mtb [51]. Although microarray has produced a significant amount of transcriptional expression information [42, 52], proteomics allows the study of final gene products in action and their interaction with the microenvironment they are in [53]. The

label-free quantitative proteomics system allowed us to unbiasedly investigate the response of Msm to an acidic growth condition. We also incorporated a labeled internal standard sample to estimate and control the false discovery rate when the number of analysis replicates were limited [13].

Out of the 1032 detected proteins, the acid stress induced 21 proteins (**Table 1**). It is noteworthy that 14 of the 20 Mtb genes homologous to these upregulated Msm proteins were most significantly induced at 24 hour after Mtb infected the activated macrophage (**Table 3**). For convenience, the 20 Mtb genes homologous to the upregulated Msm proteins are called "the 20 homologous Mtb genes" hereafter. Oxidative stress introduced with the addition of hydrogen peroxide to the Mtb 7H9 broth culture did not induce these 20 homologous Mtb genes except for *fadD9* (Rv2590) which is a homologue of MSMEG_2956 [42]. The lack of an effect of hydrogen peroxide on 19 of the 20 homologous Mtb genes suggests that oxidative stress was not be the only factor to induce 14 of the 20 homologous Mtb genes. The presence of palmitic acid in the Mtb 7H9 broth culture induced none of these 20 homologous Mtb genes, excluding that fatty acids in macrophages would induce 14 of the 20 homologous Mtb genes in activated macrophages. Interestingly, only three of the 20 homologous Mtb genes were marginally induced in naïve macrophages (**Table 3**). It appears that the induction of the 20 homologous Mtb genes was specific to activated macrophages with a peak response at 24 hours post infection. Macrophage activation leads to phagosome maturation and acidification. These results suggest that acid stress response could play a role in Mtb pathogenesis.

One of the mycobacterial virulence factors, the alpha-crystallin-like protein (*acr*; Rv2031c and MSMEG_3932), was most significantly induced both in the intraphagosomal Mtb in activated macrophages and in the acid stressed Msm cells in 7H9 culture (**Tables 1** & **3**). The alpha-crystallin-like protein, or the 14-kDa antigen, was the third most significantly induced gene (by 17-fold) among the 20 homologous Mtb genes, and is the most significantly induced protein (by 12-fold) among the 21 upregulated Msm proteins. This result indicates that acid stress could induce some of the mycobacterial virulence factors and further supports that acidic adaptation involved in Mtb pathogenesis.

We performed a gene ontology analysis for the 52 differentially regulated proteins (**Table 1**) to assess their enrichment among the detected proteins with the bioinformatics software Cytoscape [35]. The Gene Ontology (GO) project provides controlled vocabularies for the description of biological process, molecular function, and cellular component in a cell [32]. The GO terms can be used as attributes of proteins to facilitate uniform queries. Although GO does not necessarily represent the best functional categorization of genes for all occasions because redundancy and interdependence exist in the GO network, we choose to perform GO analysis of the 52 differentially regulated proteins in this study for several reasons. First, we try not to rely only on the one-gene-one-genotype paradigm to interpret the roles of the differentially regulated proteins in the acid stressed Msm cells. Second, GO categorizes proteins into sets represented by the GO terms. These protein sets allow more rigorous statistical analysis to improve confidence in a biological conclusion. For example, the presence of multiple differentially regulated proteins in a biological process indicates that the biological process is more likely to be differentially regulated than when only a single differentially regulated protein is present in that biological process. Third, GO provides a platform to evaluate the influence of the acid stress from three complementary ontologies i.e., biological process, cellular component, and molecular function. Such a sub-proteomic analysis is useful when a particular ontology needs to be analyzed, for example, to gain information of enrichment of proteins in specific organelles [17, 54] or subcellular localization [55]. Thus, a GO analysis provides a system-based approach to interpret large scale proteomics data [56], focuses on the behavior of protein sets, and thus bears a higher statistical confidence in a biological conclusion compared to an analysis of individual proteins alone.

The gene ontology analysis revealed that two major biological processes were affected by the acid stress. One is membrane transport activity and the other is fatty acid metabolism. In a DNA microarray analysis of the global transcriptional response of Mtb to a low pH under *in vitro* conditions, Fisher et al. identified 81 genes differentially expressed at >1.5-

fold in a pH 5.5 versus pH 6.9 growth medium [39]. Many of those genes involved in fatty acid metabolism, consistent with our finding that proteins involved in fatty acid metabolism were upregulated in the acid stressed Msm cells.

Upon infection, mycobacteria sense the intraphagosomal environment and synthesize critical proteins to interfere with the host defense. Although Msm is non-pathogenic, it possesses the ability to initially delay the acidification of the phagosome at the early stage of entry into macrophages [12, 57]. This ability of Msm to delay acidification of phagosomes is similar to that of Mtb. Although the ability of Mtb to resist phagolysosome fusion is a hallmark in its pathogenesis, Mtb can survive in acidified compartments. Mtb infection of freshly isolated human alveolar macrophages revealed that the majority of phagosomes containing intact bacteria were in an intimate contact with lysosomes [58]. 'Turning on' the phagolysosome fusion or 'reversing' the usual non-fusion pattern in normal mouse peritoneal macrophages did not influence the outcome of infection with virulent Mtb. Lysosome contents manifestly failed to exercise an antibacterial effect on Mtb [59]. The failure of the lysosome contents to kill Mtb after its adaptation to an intracellular environment suggests that an initial acid stress response prepares the mycobacterium for subsequent survival in the hostile intracellular environment. Indeed, it was suggested that the fusion of Mtb-containing phagosomes with lysosomes at the initial stage of infection was actually advantageous for the survival of the Mtb cells in the macrophage [4]. It seems quite likely that the bacteria have adapted to be able to live within a phagolysosome in activated macrophages [60-62].

Thus, mycobacteria possibly have at least two sets of adaptive mechanisms [61]. One is to restrict the phagolysosome fusion early in infection. The other is to adapt to phagolysosome fusion within the activated macrophages of granulomas later in infection. In a prior study, we showed that Msm could readily adapt to an acid shock by significantly readjusting its proteome to resume growth at an acidic condition [21]. The results from this study suggest that reduced membrane transport could be part of the acidic adaptation process. Inadvertently, Purdy et al. recently showed that decreased outer membrane permeability protects Msm from killing by the bactericidal action of ubiquitin-derived peptides and macrophages [10].

The induction of fatty acid metabolism genes or proteins by acid stress might contribute to the thickening of the cell wall to protect the cells or to the storage of an energy source when the cells transit to a slower growth or non-replicating state under the stress. It is not clear, however, whether the upregulated fatty acid metabolism proteins could also relate to accelerated fatty acid catabolism under acid stress, which probably requires a metabolite analysis to confirm.

Conclusion

We found that the most significant changes induced by a low pH include a down-regulation of proteins involved in transmembrane transporter activity and an upregulation of enzymes involved in fatty acid metabolism based on a GO analysis. While an inspection of other differentially regulated proteins individually could also lead to conclusions about the involvement of those proteins in the acid stress response, the GO enrichment analyses reduce the possibility that some proteins may appear to be involved in the stress response merely out of chance. The functions of many Msm proteins are annotated based on homology and not yet functionally confirmed [63]. There is also a possibility that the upregulation of some proteins is due to regulon overlap or pathway crosstalk instead of a function specific to the particular stress [64]. Thus, before further functional analysis can be carried out for many proteins, we use the GO analysis approach to distill the major changes that occurred in the acid stressed Msm cells. The GO analysis should lead to a more reliable conclusion than an interpretation of individual proteins alone.

On the other hand, GO might not capture all of the significant changes. For example, a protein interaction network analysis revealed the upregulation of a 3-protein cluster including MSMEG_5243, MSMEG_5245, and MSMEG_5246, which are related to the DevSR two-component system. Thus, an alternative gene interaction network analysis could complement the GO analysis [65]. Although further analysis of the role of the Msm DevSR two-component system in acid stress is out of the scope of this study, the upregulation of these

three proteins implies that the Msm DevSR two-component system probably responds to a broader range of stresses in addition to heat, hypoxia, and NO [11, 23].

Acknowledgement

We thank Giovanni Lostumbo and Prahlad Rao for proofreading the manuscript.

Address correspondence to: Qingbo Li, PhD, Center for Pharmaceutical Biotechnology, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60607, USA; Department of Microbiology and Immunology, College of Medicine, University of Illinois at Chicago, Chicago, IL 60612, USA; Tel: 312-413-9301; Fax: 312-413-9303, E-mail: qkli@uic.edu

References

[1] Hestvik AL, Hmama Z and Av-Gay Y. Mycobacterial manipulation of the host cell. FEMS Microbiol Rev 2005; 29: 1041-1050.

[2] Crowle AJ, Dahl R, Ross E and May MH. Evidence that vesicles containing living, virulent Mycobacterium tuberculosis or Mycobacterium avium in cultured human macrophages are not acidic. Infect Immun 1991; 59: 1823-1831.

[3] Sturgill-Koszycki S, Schlesinger PH, Chakraborty P, Haddix PL, Collins HL, Fok AK, Allen RD, Gluck SL, Heuser J and Russell DG. Lack of acidification in Mycobacterium phagosomes produced by exclusion of the vesicular proton-ATPase. Science 1994; 263: 678-681.

[4] Brown CA, Draper P and Hart PD. Mycobacteria and lysosomes: a paradox. Nature 1969; 221: 658-660.

[5] Vandal OH, Pierini LM, Schnappinger D, Nathan CF and Ehrt S. A membrane protein preserves intrabacterial pH in intraphagosomal Mycobacterium tuberculosis. Nat Med 2008; 14: 849-854.

[6] McDonough KA, Kress Y and Bloom BR. Pathogenesis of tuberculosis: interaction of Mycobacterium tuberculosis with macrophages. Infect Immun 1993; 61: 2763-2773.

[7] McDonough KA, Florczyk MA and Kress Y. Intracellular passage within macrophages affects the trafficking of virulent tubercle bacilli upon reinfection of other macrophages in a serum-dependent manner. Tuber Lung Dis 2000; 80: 259-271.

[8] Singh AK and Reyrat JM. Laboratory maintenance of Mycobacterium smegmatis. Curr Protoc Microbiol 2009; Chapter 10: Unit10C 11.

[9] Jhamb SS and Singh PP. A short-term model for preliminary screening of potential anti-tubercular compounds. Scand J Infect Dis 2009; 1-4.

[10] Purdy GE, Niederweis M and Russell DG. Decreased outer membrane permeability protects mycobacteria from killing by ubiquitin-derived peptides. Mol Microbiol 2009; 73: 844-857.

[11] Lee JM, Cho HY, Cho HJ, Ko IJ, Park SW, Baik HS, Oh JH, Eom CY, Kim YM, Kang BS and Oh JI. O2- and NO-sensing mechanism through the DevSR two-component system in Mycobacterium smegmatis. J Bacteriol 2008; 190: 6795-6804.

[12] Anes E, Peyron P, Staali L, Jordao L, Gutierrez MG, Kress H, Hagedorn M, Maridonneau-Parini I, Skinner MA, Wildeman AG, Kalamidas SA, Kuehnel M and Griffiths G. Dynamic life and death interactions between Mycobacterium smegmatis and J774 macrophages. Cell Microbiol 2006; 8: 939-960.

[13] Li Q and Roxas BA. An assessment of false discovery rates and statistical significance in label-free quantitative proteomics with combined filters. BMC Bioinformatics 2009; 10: 43.

[14] Roxas BA and Li Q. Significance analysis of microarray for relative quantitation of LC/MS data in proteomics. BMC Bioinformatics 2008; 9: 187.

[15] Rao PK, Rodriguez GM, Smith I and Li Q. Protein dynamics in iron-starved Mycobacterium tuberculosis revealed by turnover and abundance measurement using hybrid-linear ion trap-fourier transform mass spectrometry. Anal Chem 2008; 80: 6860-6869.

[16] Andreev VP, Li L, Cao L, Gu Y, Rejtar T, Wu SL and Karger BL. A new algorithm using cross-assignment for label-free quantitation with LC-LTQ-FT MS. J Proteome Res 2007; 6: 2186-2194.

[17] Rao PK, Singh CR, Jagannath C and Li Q. A systems biology approach to study the phagosomal proteome modulated by mycobacterial infections. Int J Clin Exp Med 2009; 2: 233-247.

[18] Katti MK, Dai G, Armitige LY, Rivera Marrero C, Daniel S, Singh CR, Lindsey DR, Dhandayuthapani S, Hunter RL and Jagannath C. The Delta fbpA mutant derived from Mycobacterium tuberculosis H37Rv has an enhanced susceptibility to intracellular antimicrobial oxidative mechanisms, undergoes limited phagosome maturation and activates macrophages and dendritic cells. Cell Microbiol 2008; 10: 1286-1303.

[19] Geisow MJ, D'Arcy Hart P and Young MR. Temporal changes of lysosome and phagosome pH during phagolysosome formation in macrophages: studies by fluorescence spectroscopy. J Cell Biol 1981; 89: 645-652.

[20] Pethe K, Swenson DL, Alonso S, Anderson J,

Wang C and Russell DG. Isolation of Mycobacterium tuberculosis mutants defective in the arrest of phagosome maturation. Proc Natl Acad Sci U S A 2004; 101: 13642-13647.
[21] Rao PK, Roxas BA and Li Q. Determination of global protein turnover in stressed mycobacterium cells using hybrid-linear ion trap-fourier transform mass spectrometry. Anal Chem 2008; 80: 396-406.
[22] http://string.embl.de/.
[23] O'Toole R, Smeulders MJ, Blokpoel MC, Kay EJ, Lougheed K and Williams HD. A two-component regulator of universal stress protein expression and adaptation to oxygen starvation in Mycobacterium smegmatis. J Bacteriol 2003; 185: 1543-1554.
[24] Roback P, Beard J, Baumann D, Gille C, Henry K, Krohn S, Wiste H, Voskuil MI, Rainville C and Rutherford R. A predicted operon map for Mycobacterium tuberculosis. Nucleic Acids Res 2007; 35: 5085-5095.
[25] Matsoso LG, Kana BD, Crellin PK, Lea-Smith DJ, Pelosi A, Powell D, Dawes SS, Rubin H, Coppel RL and Mizrahi V. Function of the cytochrome bc1-aa3 branch of the respiratory network in mycobacteria and network adaptation occurring in response to its disruption. J Bacteriol 2005; 187: 6300-6308.
[26] Mayuri, Bagchi G, Das TK and Tyagi JS. Molecular analysis of the dormancy response in Mycobacterium smegmatis: expression analysis of genes encoding the DevR-DevS two-component system, Rv3134c and chaperone alpha-crystallin homologues. FEMS Microbiol Lett 2002; 211: 231-237.
[27] Sinha S, Rappu P, Lange SC, Mantsala P, Zalkin H and Smith JL. Crystal structure of Bacillus subtilis YabJ, a purine regulatory protein and member of the highly conserved YjgF family. Proc Natl Acad Sci U S A 1999; 96: 13074-13079.
[28] Dietrich LE, Teal TK, Price-Whelan A and Newman DK. Redox-active antibiotics control gene expression and community behavior in divergent bacteria. Science 2008; 321: 1203-1206.
[29] Cundliffe E. How antibiotic-producing organisms avoid suicide. Annu Rev Microbiol 1989; 43: 207-233.
[30] Morishita R, Kawagoshi A, Sawasaki T, Madin K, Ogasawara T, Oka T and Endo Y. Ribonuclease activity of rat liver perchloric acid-soluble protein, a potent inhibitor of protein synthesis. J Biol Chem 1999; 274: 20688-20692.
[31] Cashman JR. Role of flavin-containing monooxygenase in drug development. Expert Opin Drug Metab Toxicol 2008; 4: 1507-1521.
[32] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM and Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000; 25: 25-29.
[33] ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/.
[34] Maere S, Heymans K and Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 2005; 21: 3448-3449.
[35] http://www.cytoscape.org/.
[36] Postma PW, Lengeler JW and Jacobson GR. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol Rev 1993; 57: 543-594.
[37] Kights KM. Long-Chain-Fatty-Acid CoA Ligases: The Key to Fatty Acid Activation, Formation of Xenobiotic Acyl-CoA Thioesters and Lipophilic Xenobiotic Conjugates. Current Medicinal Chemistry - Immunology, Endocrine & Metabolic Agents 2003; 3: 235-244.
[38] Deb C, Lee CM, Dubey VS, Daniel J, Abomoelak B, Sirakova TD, Pawar S, Rogers L and Kolattukudy PE. A novel in vitro multiple-stress dormancy model for Mycobacterium tuberculosis generates a lipid-loaded, drug-tolerant, dormant pathogen. PLoS One 2009; 4: e6077.
[39] Fisher MA, Plikaytis BB and Shinnick TM. Microarray analysis of the Mycobacterium tuberculosis transcriptional response to the acidic conditions found in phagosomes. J Bacteriol 2002; 184: 4025-4032.
[40] Xia Q, Hendrickson EL, Zhang Y, Wang T, Taub F, Moore BC, Porat I, Whitman WB, Hackett M and Leigh JA. Quantitative proteomics of the archaeon Methanococcus maripaludis validated by microarray analysis and real time PCR. Mol Cell Proteomics 2006; 5: 868-881.
[41] Li Q. Advances in protein turnover analysis at the global level and biological insights. Mass Spectrom Rev 2009; Sep 15 [Epub ahead of print].
[42] Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C and Schoolnik GK. Transcriptional Adaptation of Mycobacterium tuberculosis within Macrophages: Insights into the Phagosomal Environment. J Exp Med 2003; 198: 693-704.
[43] Gillespie E. Effects of S-lactoylglutathione and inhibitors of glyoxalase I on histamine release from human leukocytes. Nature 1979; 277: 135-137.
[44] Kim NS, Umezawa Y, Ohmura S and Kato S. Human glyoxalase I. cDNA cloning, expression, and sequence similarity to glyoxalase I from Pseudomonas putida. J Biol Chem 1993; 268: 11217-11221.
[45] Thornalley PJ. Glyoxalase I--structure, function and a critical role in the enzymatic defence against glycation. Biochem Soc Trans 2003; 31: 1343-1348.
[46] Hao da C, Chen SL, Mu J and Xiao PG.

Molecular phylogeny, long-term evolution, and functional divergence of flavin-containing monooxygenases. Genetica 2009; 137: 173-187.
[47] Betts JC, Lukey PT, Robb LC, McAdam RA and Duncan K. Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. Mol Microbiol 2002; 43: 717-731.
[48] Qian L and Ortiz de Montellano PR. Oxidative activation of thiacetazone by the Mycobacterium tuberculosis flavin monooxygenase EtaA and human FMO1 and FMO3. Chem Res Toxicol 2006; 19: 443-449.
[49] http://www.geneontology.org/.
[50] http://www.ncbi.nlm.nih.gov/Structure/index.shtml.
[51] Dick T, Lee BH and Murugasu-Oei B. Oxygen depletion induced dormancy in Mycobacterium smegmatis. FEMS Microbiol Lett 1998; 163: 159-164.
[52] Hampshire T, Soneji S, Bacon J, James BW, Hinds J, Laing K, Stabler RA, Marsh PD and Butcher PD. Stationary phase gene expression of Mycobacterium tuberculosis following a progressive nutrient depletion: a model for persistent organisms? Tuberculosis (Edinb) 2004; 84: 228-238.
[53] Nissom PM, Sanny A, Kok YJ, Hiang YT, Chuah SH, Shing TK, Lee YY, Wong KT, Hu WS, Sim MY and Philp R. Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells. Mol Biotechnol 2006; 34: 125-140.
[54] Trost M, English L, Lemieux S, Courcelles M, Desjardins M and Thibault P. The phagosomal proteome in interferon-gamma-activated macrophages. Immunity 2009; 30: 143-154.
[55] Dumas E, Desvaux M, Chambon C and Hebraud M. Insight into the core and variant exoproteomes of Listeria monocytogenes species by comparative subproteomic analysis. Proteomics 2009; 9: 3136-3155.
[56] Feltrin E, Campanaro S, Diehl AD, Ehler E, Faulkner G, Fordham J, Gardin C, Harris M, Hill D, Knoell R, Laveder P, Mittempergher L, Nori A, Reggiani C, Sorrentino V, Volpe P, Zara I, Valle G and Deegan Nee Clark J. Muscle Research and Gene Ontology: New standards for improved data integration. BMC Med Genomics 2009; 2: 6.
[57] Kuehnel MP, Goethe R, Habermann A, Mueller E, Rohde M, Griffiths G and Valentin-Weigand P. Characterization of the intracellular survival of Mycobacterium avium ssp. paratuberculosis: phagosomal pH and fusogenicity in J774 macrophages compared with other mycobacteria. Cell Microbiol 2001; 3: 551-566.
[58] Borelli V, Vita F, Soranzo MR, Banfi E and Zabucchi G. Ultrastructure of the interaction between mycobacterium tuberculosis- H37Rv-containing phagosomes and the lysosomal compartment in human alveolar macrophages. Exp Mol Pathol 2002; 73: 128-134.
[59] Armstrong JA and Hart PD. Phagosome-lysosome interactions in cultured macrophages infected with virulent tubercle bacilli. Reversal of the usual nonfusion pattern and observations on bacterial survival. J Exp Med 1975; 142: 1-16.
[60] Chan K, Knaak T, Satkamp L, Humbert O, Falkow S and Ramakrishnan L. Complex pattern of Mycobacterium marinum gene expression during long-term granulomatous infection. Proc Natl Acad Sci U S A 2002; 99: 3920-3925.
[61] Cosma CL, Sherman DR and Ramakrishnan L. The secret lives of the pathogenic mycobacteria. Annu Rev Microbiol 2003; 57: 641-676.
[62] Ramakrishnan L, Federspiel NA and Falkow S. Granuloma-specific expression of Mycobacterium virulence proteins from the glycine-rich PE-PGRS family. Science 2000; 288: 1436-1439.
[63] Wang R, Prince JT and Marcotte EM. Mass spectrometry of the M. smegmatis proteome: protein expression levels correlate with function, operons, and codon bias. Genome Res 2005; 15: 1118-1126.
[64] Pomposiello PJ, Bennik MH and Demple B. Genome-wide transcriptional profiling of the Escherichia coli responses to superoxide stress and sodium salicylate. J Bacteriol 2001; 183: 3890-3902.
[65] Balazsi G, Heath AP, Shi L and Gennaro ML. The temporal response of the Mycobacterium tuberculosis gene regulatory network during growth arrest. Mol Syst Biol 2008; 4: 225.