

**Top-down proteomics reveals a unique protein S-thiolation switch in *Salmonella*
Typimurium in response to infection-like conditions**

Charles Ansong^{1*#}, Si Wu^{2*}, Da Meng³, Xiaowen Liu⁴, Heather Brewer², Brooke L. Deatherage
Kaiser¹, Ernesto S. Nakayasu¹, John R. Cort¹, Pavel A. Pevzner⁵, Richard D. Smith¹, Fred
Heffron⁶, Joshua N. Adkins^{1#} and Ljilijana Pasa-Tolic^{2#}

* Contributed equally

Corresponding authors

¹Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, 99352 USA

²Environmental Molecular Science Laboratory, Pacific Northwest National Laboratory, Richland, WA, 99352 USA

³Computational Science and Mathematics Division, Pacific Northwest National Laboratory, Richland, WA, 99352
USA

⁴School of Informatics, Indiana University-Purdue University Indianapolis, Indianapolis, IN 46202 USA

⁵Department of Computer Science and Engineering, University of California San Diego, San Diego, CA, 92093,
USA

⁶Department of Microbiology and Immunology, Oregon Health and Science University, Portland, OR, 97239, USA

Supplementary Methods

Media and Chemical reagents

Bacterial cells were grown in Luria-Bertani (LB) medium (Difco, Franklin Lakes, NJ, USA) or low phosphate low magnesium (LPM) acidic minimal medium as described previously (1). Briefly, LPM contained 5 mM KCl, 7.5 mM (NH₄)₂SO₄, 0.5 mM K₂SO₄, 0.3% (vol/vol) glycerol, 0.00001% thiamine, 0.5 μM ferric citrate, 8 μM MgCl₂, 337 μM PO₄³⁻, and 80 mM MES [2-(N-morpholino)ethanesulfonic acid]-free acid adjusted to pH 5.8 with NaOH. Components of LPM (1) and other chemicals were purchased from Sigma (St. Louis, MO, USA). Protein concentrations were determined by bicinchoninic acid (BCA) assay (Pierce, Rockford, IL, USA) using bovine serum albumin as standards. Trypsin used for protein digestions was purchased from Promega (Madison, WI, USA). Slide-a-lyzer dialysis device was purchased from Pierce (Rockford, IL, USA).

Bottom-up Sample preparation

Overnight starter cultures of *Salmonella* Typhimurium 14028 were grown in LB broth at 37°C with shaking at 300rpm. The cultures were diluted 1:100 into LB and grown to mid-log phase (OD) for the basal LB condition or diluted 1:10 into LPM and grown for 20h for infection-like LPM condition respectively as previously described (1, 2). Cell pellets were re-suspended in 100 mM NH₄HCO₃, pH 8.4 buffer and lysed by using 0.1 mM zirconia/silica beads in a 2.0-ml Cryovial with vigorous vortexing for a total of 3 min with cooling steps. The supernatant and subsequent washes were transferred from the beads into new Cryovials. The beads were repeatedly washed until the supernatant was clear. After protein concentration was determined for the samples, urea and thiourea were added to final concentrations of 7 and 2 M, respectively. The samples were incubated at 60°C for 30 min. and then diluted 10-fold with buffer, and CaCl₂ was added (1 mM) followed by trypsin in a 1:50 trypsin:protein ratio. The samples were digested for 3 h at 37°C and subsequently cleaned using a C18 solid phase extraction (SPE) column (Supelco). Each 1-ml 100- or 50-mg SPE column was conditioned with methanol and rinsed with 0.1% trifluoroacetic acid (TFA) in water. Samples were introduced to the columns and then washed with 95:5 H₂O:acetonitrile (ACN) that contained 0.1% TFA. Excess liquid was removed from the columns under vacuum, and the samples were eluted with 80:20 ACN:H₂O that contained 0.1% TFA and concentrated in a SpeedVac (Thermo-Savant) to a final volume of

~100 μ l. A BCA protein assay was performed to determine peptide concentrations prior to analysis.

Top-down Sample preparation

Overnight starter cultures of *Salmonella* Typhimurium 14028 were grown in LB broth at 37°C with shaking at 300 rpm. The cultures were diluted 1:100 into LB and grown to mid-log phase (OD) for the basal LB condition or diluted 1:10 into LPM and grown for 20h for infection-like LPM condition respectively as previously described (1, 2). Cell pellets were resuspended in 100 mM NH_4HCO_3 , pH 8.4 buffer with 1mM PMSF protease inhibitor and lysed by using 0.1 mM zirconia/silica beads in a 2.0-ml Cryovial with vigorous vortexing for a total of 3 min with cooling steps. The supernatant and subsequent washes were transferred from the beads into new Cryovials. The beads were repeatedly washed until the supernatant was clear. After protein concentration was determined for the samples, urea was added to final concentrations of 7 M. Following dissolution of the urea, the samples were incubated at 60°C for 30 min. Samples were then ultracentrifuged at 100,000 rpm for 10 minutes at 4 °C. The supernatant were transferred into a Slide-a-lyzer using a syringe and dialyzed extensively overnight with at least 3 buffer changes.

Bottom-up LC-MS/MS Analysis

The high-performance liquid chromatography (HPLC) system and method used for nanocapillary liquid chromatography have been described in detail elsewhere (3-6). Analysis was performed using an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) with electrospray ionization. The HPLC column was coupled to the mass spectrometer using an in-house-manufactured interface. The heated capillary temperature and spray voltage were 200°C and 2.2 kV, respectively. Data acquisition began 20 min after the sample was injected and continued for 100 min over an m/z range of 400 to 2,000. For each cycle, the six most abundant ions from MS analysis were selected for MS/MS analysis, using a collision energy of 35% A dynamic exclusion time of 60 s was used to discriminate against previously analyzed ions. Each sample was analyzed in duplicate.

Top-down LC-MS/MS Analysis

The intact protein RPLC separation was performed on a Waters NanoAquity mass spectrometer with a column (80 cm $\times$ 75 μ m i.d.) packed in-house with Phenomenex Jupiter particles (C5

stationary phase, 5 μm particle diameter, 300 $\text{\AA}$ pore size). Mobile phase A was composed of 0.5% acetic acid, 0.01% TFA, 5% isopropanol, 10% ACN, and 84.5% water. Mobile phase B consisted of 0.5% acetic acid, 0.01% TFA, 9.9% water, 45% isopropanol, and 45% ACN. The operating flow rate was 0.3 $\mu\text{l}/\text{min}$. The RPLC system was equilibrated with 100% mobile phase A for 5 minutes, and then increased to 20% mobile phase B in 1 minute. A 250 minute linear gradient was set from 20% mobile phase B to 55% mobile phase B. MS analysis was performed using an LTQ Orbitrap Velos ETD mass spectrometer (Thermo Scientific, San Jose, CA) outfitted with a custom electrospray ionization (ESI) interface. Electrospray emitters were custom made using 150 μm o.d. x 20 μm i.d. chemically etched fused silica (7). The heated capillary temperature and spray voltage were 275°C and 2.2 kV, respectively. For the LC-MS/MS analysis with ETD fragmentation, a parent spectrum was collected at a 60K resolution and was followed by high resolution (60K) ETD tandem mass spectra of the 8 most intense ions from the parent mass spectrum. The ETD reaction time was fixed at 40 ms. For the LC-MS/MS analysis with HCD fragmentation, a parent spectrum was collected at a 60K resolution and was followed by high resolution (60K) HCD tandem mass spectra of the 8 most intense ions from the parent mass spectrum. FT MS/MS employed 45% normalized collision energy for HCD. Mass calibration was performed prior to analysis according to the method recommended by the instrument manufacturer.

Bottom-up Data Analysis

Peptides were identified by using SEQUEST to search the mass spectra from LC-MS/MS analyses. These searches were performed using the annotated Typhimurium 14028 FASTA file, containing 5590 protein sequences (8). Porcine trypsin protein sequences were included in the search to detect trypsin autocleavage contaminants. The SEQUEST parameter file allowed for dynamic oxidation of methionine and S-glutathionylation or S-Cysteinylation modifications to amino acid residues and peptide_mass_tolerance = 3 and fragment_ion_tolerance = 0. The searches allowed for all possible peptide termini, that is, not limited by tryptic terminus state. Peptide identifications were filtered using an MS-GF cut-off value of $1\text{E}-10$ (9) in addition to a 10 ppm precursor mass measurement error requirement. This corresponded to a false discovery rate of <1% at the unique peptide level. Spectral count (10, 11) was used as one method for quantification. Additionally we also employed an intensity based quantification approach. Briefly, all the identified peptides with modifications were combined to a candidate list. For

each peptide on this list, the theoretical mass was calculated based on the sequence information. Observed charge states from the identifications were also included in the list to facilitate the calculation of theoretical masses of each possible m/z ion species. A targeted peak extraction approach similar to the AMT tag approach (12) was then applied against each charge state using a 15 ppm window. For each peptide, all the peak areas corresponding to different charge states (with the correct elution mass and retention time) were then summed for final intensity comparison. Missing values were imputed with the row mean within each factor where 2 or more values are present or with the minimal dataset value where less than 2 values are present. Differences in log-transformed peptide intensity/abundance between LB and LPM growth conditions was assessed via hypothesis testing, specifically a comprehensive ANOVA scheme included in the software program DANTE (13) and fold changes calculated.

Top-down Data Analysis

To obtain protein identifications top-down mass spectra were analyzed using the MS-Align+ software as previously described (14). The database used for searching was the annotated *Salmonella* Typhimurium 14028 FASTA file, containing 5590 protein sequences (8). False discovery rate (FDR) was estimated as previously described (15, 16), and results from target-decoy searches are shown in Supplementary Figure 17. The “bird eye view” based relative quantitation was done by summing all the spectra under the corresponding LC peaks. The resulting spectrum was reprocessed by collapsing all charge states in the detected m/z range into a zero charge state spectrum using in-house developed software ICR-2LS. The theoretical spectrum for each corresponding protein was also generated using ICR-2LS based on the accurate molecular formula with 100k assigned resolution.

References

1. Niemann GS, *et al.* (2011) Discovery of novel secreted virulence factors from *Salmonella enterica* serovar Typhimurium by proteomic analysis of culture supernatants. *Infect Immun* 79(1):33-43 .
2. Brown RN, *et al.* (2012) A Comprehensive Subcellular Proteomic Survey of *Salmonella* Grown under Phagosome-Mimicking versus Standard Laboratory Conditions. *Int J Proteomics* 2012:123076 .
3. Adkins JN, *et al.* (2006) Analysis of the *Salmonella typhimurium* proteome through environmental response toward infectious conditions. *Mol Cell Proteomics* 5(8):1450-1461 .

4. Livesay EA, et al. (2008) Fully automated four-column capillary LC-MS system for maximizing throughput in proteomic analyses. *Anal Chem* 80(1):294-302 .
5. Shen Y, et al. (2001) High-throughput proteomics using high-efficiency multiple-capillary liquid chromatography with on-line high-performance ESI FTICR mass spectrometry. *Anal Chem* 73(13):3011-3021 .
6. Shi L, et al. (2009) Proteome of Salmonella Enterica Serotype Typhimurium Grown in a Low Mg/pH Medium. *J Proteomics Bioinform* 2:388-397 .
7. Kelly RT, Page JS, Tang K, & Smith RD (2007) Array of chemically etched fused-silica emitters for improving the sensitivity and quantitation of electrospray ionization mass spectrometry. *Anal Chem* 79(11):4192-4198 .
8. Jarvik T, Smillie C, Groisman EA, & Ochman H (2010) Short-term signatures of evolutionary change in the Salmonella enterica serovar typhimurium 14028 genome. *J Bacteriol* 192(2):560-567 .
9. Kim S, Gupta N, & Pevzner PA (2008) Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J Proteome Res* 7(8):3354-3363 .
10. Liu H, Sadygov RG, & Yates JR, 3rd (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 76(14):4193-4201 .
11. Washburn MP, Wolters D, & Yates JR, 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19(3):242-247 .
12. Zimmer JS, Monroe ME, Qian WJ, & Smith RD (2006) Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom Rev* 25(3):450-482 .
13. Polpitiya AD, et al. (2008) DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 24(13):1556-1558 .
14. Liu X, et al. (2012) Protein identification using top-down. *Mol Cell Proteomics* 11(6):M111008524 .
15. Kellie JF, et al. (2012) Robust analysis of the yeast proteome under 50 kDa by molecular-mass-based fractionation and top-down mass spectrometry. *Anal Chem* 84(1):209-215 .
16. Tran JC, et al. (2011) Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* 480(7376):254-258 .

Supplementary Figure Legends

Supplementary Figure 1. Distribution of JCVI functional categories for the entire annotated genome and the characterized proteome from the present top-down study and two prior bottom-up studies.

Supplementary Figure 2. Graph showing the increase in identifications from previous studies achieved in prokaryote (bacterial) systems and comparison to eukaryote (yeast or human) systems. References cited are indicated below each bar.

Supplementary Figure 3. Switch in modification from glutathionylation to cysteinylation in response to environmental conditions for STM14_1062. A) Representative zero charge state spectra from top-down proteomics data showing switch in S-thiolation forms in STM14_1062. Fragmentation map shows specific cysteine residue on which switch occurs. B) Estimated stoichiometry from intact protein mass spectra summed across corresponding LC peak.

Supplementary Figure 4. Switch in modification from glutathionylation to cysteinylation in response to environmental conditions for STM14_0013. A) Representative zero charge state spectra from top-down proteomics data showing switch in S-thiolation forms in STM14_0013. Fragmentation map shows specific cysteine residue on which switch occurs. Estimated stoichiometry not determined due to low relative abundance of species.

Supplementary Figure 5. Switch in modification from glutathionylation to cysteinylation in response to environmental conditions for STM14_2116. A) Representative zero charge state spectra from top-down proteomics data showing switch in S-thiolation forms in STM14_2116. Fragmentation map shows specific cysteine residue on which switch occurs. B) Estimated stoichiometry from intact protein mass spectra summed across corresponding LC peak.

Supplementary Figure 6. Switch in modification from glutathionylation to cysteinylation in response to environmental conditions for STM14_4143. A) Representative zero charge state spectra from top-down proteomics data showing switch in S-thiolation forms in STM14_4143. Fragmentation map shows specific cysteine residue on which switch occurs. Estimated stoichiometry not determined due to low relative abundance of species.

Supplementary Figure 7. Switch in modification from glutathionylation to cysteinylation in response to environmental conditions for STM14_2195. A) Representative zero charge state spectra from top-down proteomics data showing switch in S-thiolation forms in STM14_2195. Fragmentation map shows specific cysteine residue on which switch occurs. B) Estimated stoichiometry from intact protein mass spectra summed across corresponding LC peak.

Supplementary Figure 8. Switch in modification from glutathionylation to cysteinylolation in response to environmental conditions for STM14_4462. A) Representative zero charge state spectra from top-down proteomics data showing switch in S-thiolation forms in STM14_4462. Fragmentation map shows specific cysteine residue on which switch occurs. B) Estimated stoichiometry from intact protein mass spectra summed across corresponding LC peak.

Supplementary Figure 9. Switch in modification from glutathionylation to cysteinylolation in response to environmental conditions for STM14_5351. A) Representative zero charge state spectra from top-down proteomics data showing switch in S-thiolation forms in STM14_5351. Fragmentation map shows specific cysteine residue on which switch occurs. B) Estimated stoichiometry from intact protein mass spectra summed across corresponding LC peak.

Supplementary Figure 10. Switch in modification from glutathionylation to cysteinylolation in response to environmental conditions for STM14_3098. A) Representative zero charge state spectra from top-down proteomics data showing switch in S-thiolation forms in STM14_3098. Fragmentation map shows specific cysteine residue on which switch occurs. B) Estimated stoichiometry from intact protein mass spectra summed across corresponding LC peak.

Supplementary Figure 11: Eight structures (1 from *Salmonella*, 6 from *E. coli*, 1 from *N. meningitides*) where cysteinylolation and glutathionylation modifications of Cys residues (spheres) were found. The red number indicates the Naccess calculated accessibility relative to Ala-Xxx-Ala for all atoms in the Cys residue, using default VdW radii and a 1.4 Å sphere. Also indicated is the pdb id, the source organism, and (in parentheses) the one letter code for the residue at the position of the modified Cys in *Salmonella*. Additionally we have also indicated the residue number. If the modified Cys of the *Salmonella* protein is not the same amino acid residue and/or residue number, then that is indicated as well with the *Salmonella* residue number in parentheses.

Supplementary Figure 12. Overlap of top-down and bottom-up protein identifications from this study.

Supplementary Figure 13. Quantification of glutathionylated and cysteinylated peptides (proteins). Graph shows spectral count based quantification of glutathionylated and cysteinylated peptides (proteins) identified from bottom-up proteomics data under LB and LPM growth conditions.

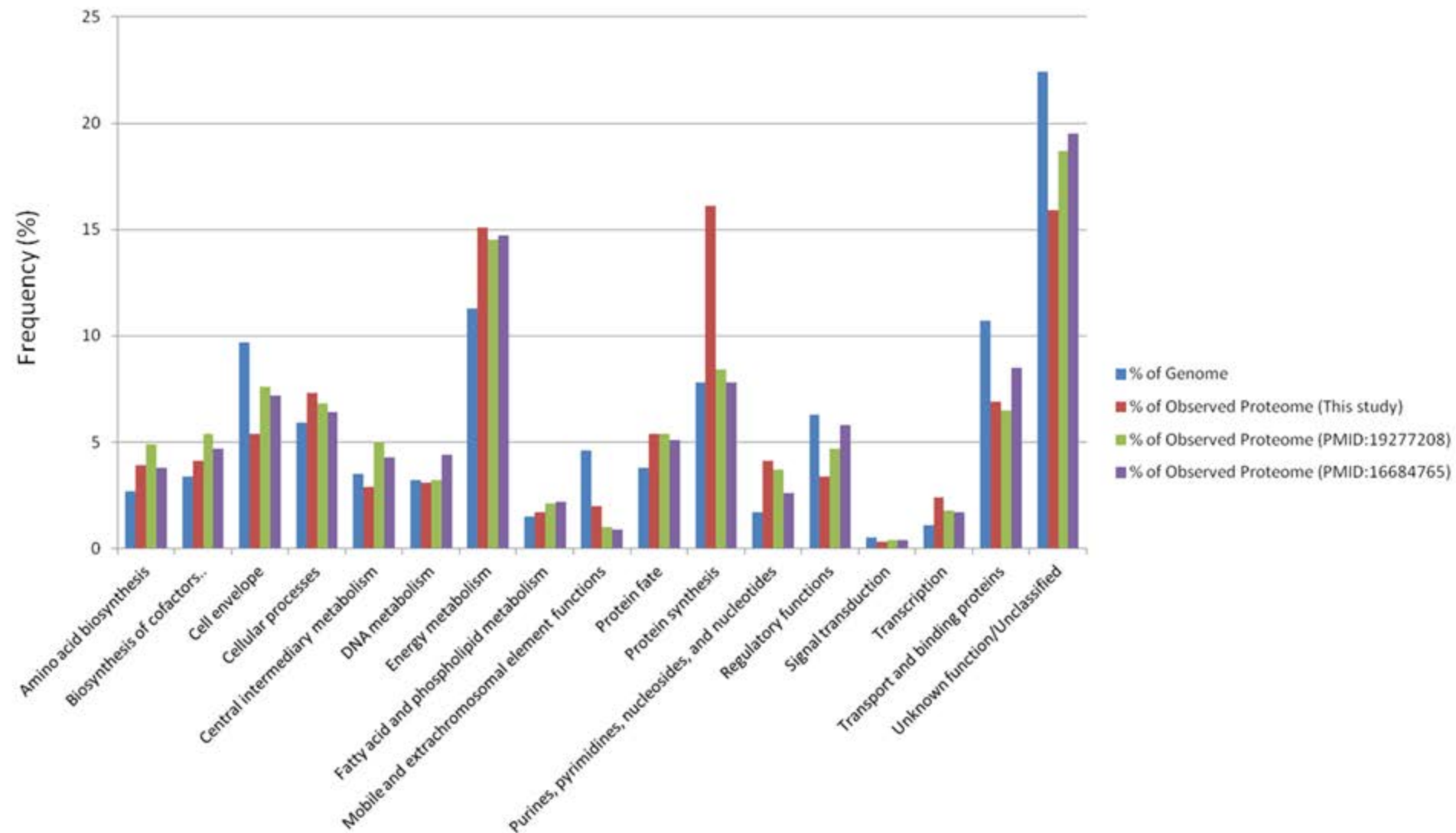
Supplementary Figure 14. Quantification of representative proteins from Figure 4 for protein abundance response to switch from LB to LPM using other (i.e. non-thiolated) peptides identified from bottom-up analysis.

Supplementary Figure 15. Quantification of glutathione and cysteine biosynthesis genes. Top panel: Fold change in protein expression of glutathione and cysteine biosynthesis genes under *in vitro* conditions. Bottom panel: Fold change in transcript expression of glutathione and cysteine biosynthesis genes under *in vivo* conditions (*Data from Hautefort et al. 2008, PMID: 18031307).

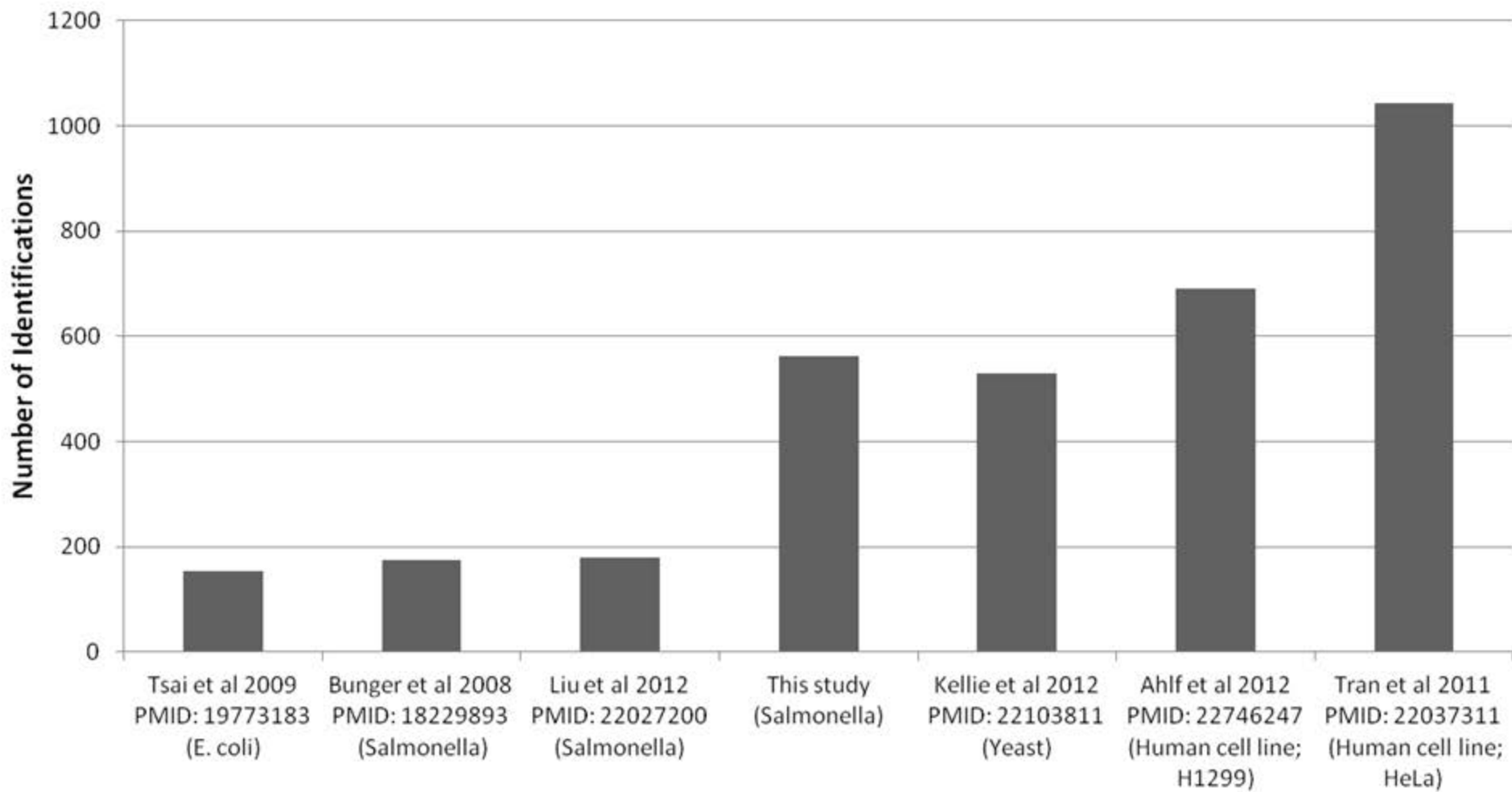
Supplementary Figure 16. Quantification of biosynthetic genes for representative non-S-containing and S-containing metabolites. Inset figure shows quantification of biosynthetic genes for representative non-S-containing (Right panel) and S-containing metabolites (Left panel) under *in vitro* conditions. Main figure shows quantification of biosynthetic genes for S-containing metabolites under *in vivo* conditions.

Supplementary Figure 17. Results from top-down target-decoy searches. P-value cut-off of $1E-4$ (1×10^{-4}) gives a cut-off for 5% FDR.

Supplementary Figure 1

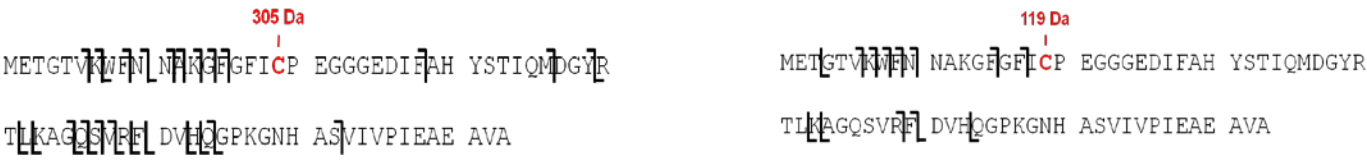
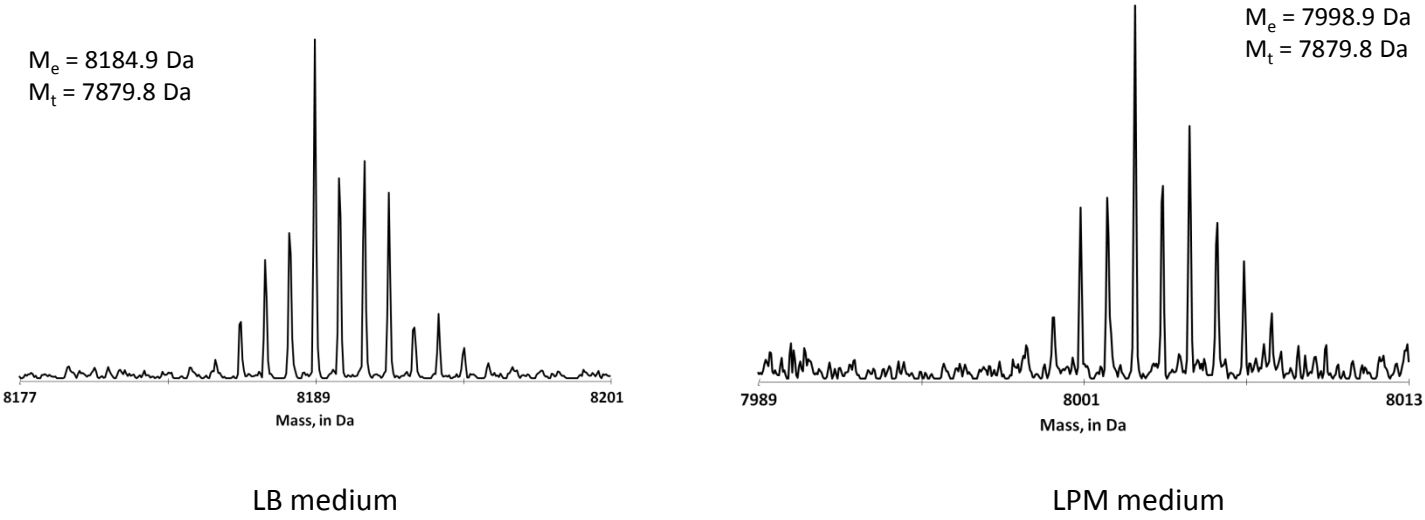


Supplementary Figure 2

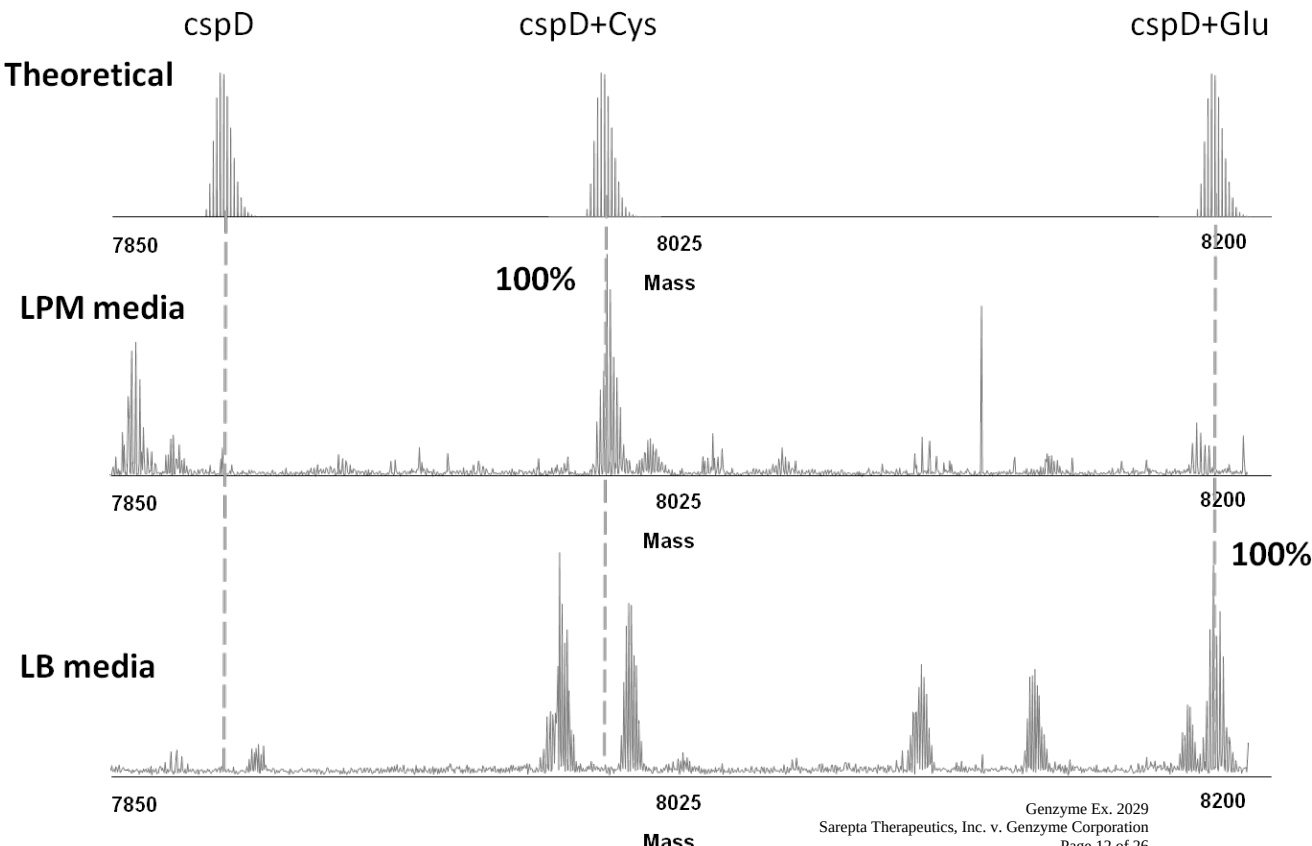


STM14_1062 cspD stress response protein

A

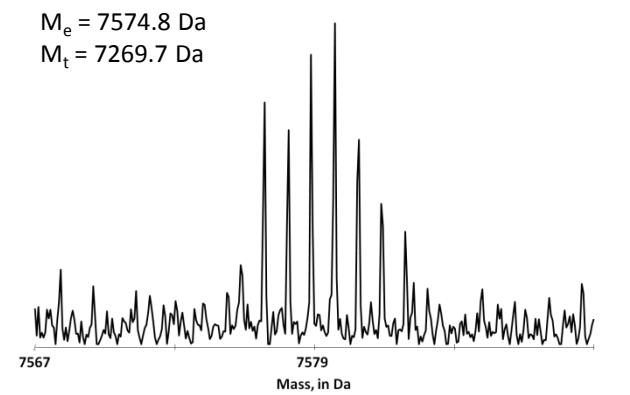


B



STM14_0013 dnaK molecular chaperone DnaK

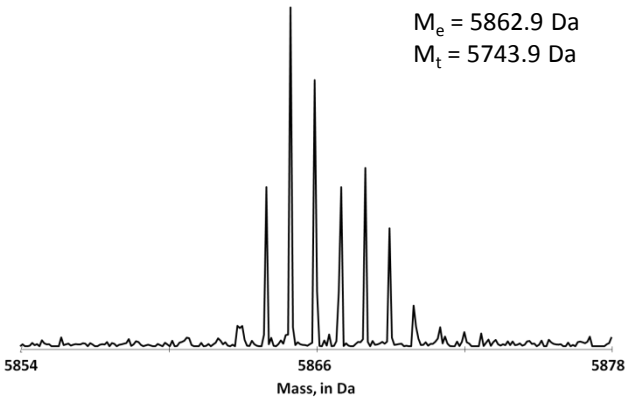
A



LB medium

305 Da

MGKILGIDLG TTNSVAIMD GTQARVLENA EGDRTTPSII
AYTIDGETLV GOPAKRQAVT NPQNTLFAIK RLIGRRFQDE
EVQRDVSIMP YKIIGADNGD AWLDVKGQKM APPQISAEVL
.....13 lines skipped.....



LPM medium

119 Da

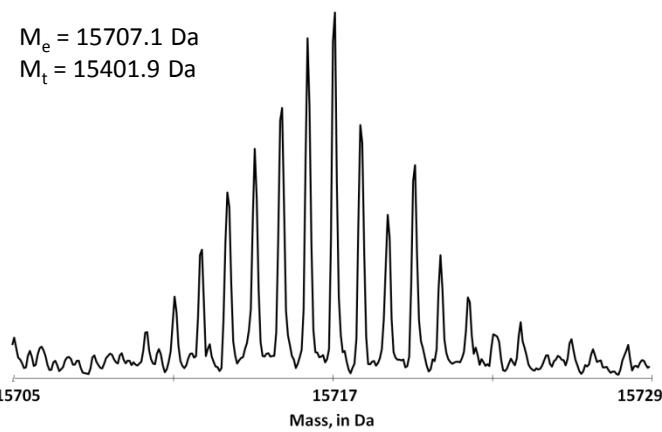
MGKILGIDLG TTNSVAIMD GTQARVLENA EGDRTTPSII
AYTIDGETLV GOPAKRQAVT NPQNTLFAIK RLIGRRFQDE
EVQRDVSIMP YKIIGADNGD AWLDVKGQKM APPQISAEVL
.....13 lines skipped.....

STM14_2116 hns global DNA-binding transcriptional dual regulator H-NS

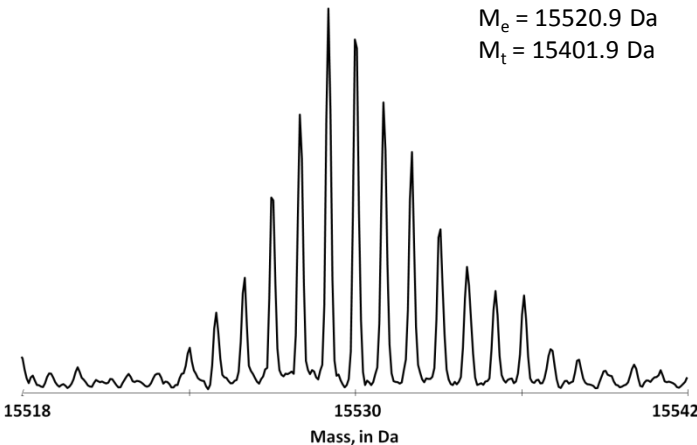
A

$M_e = 15707.1$ Da
 $M_t = 15401.9$ Da

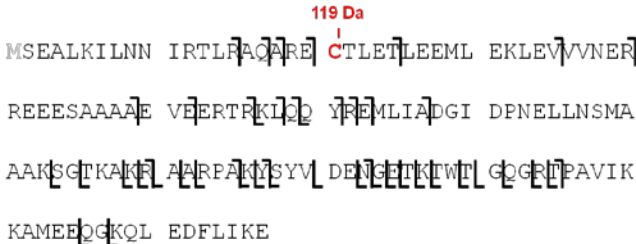
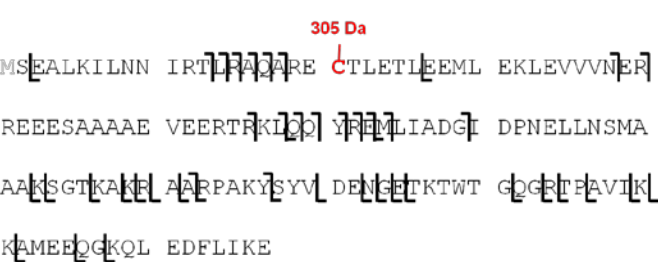
$M_e = 15520.9$ Da
 $M_t = 15401.9$ Da



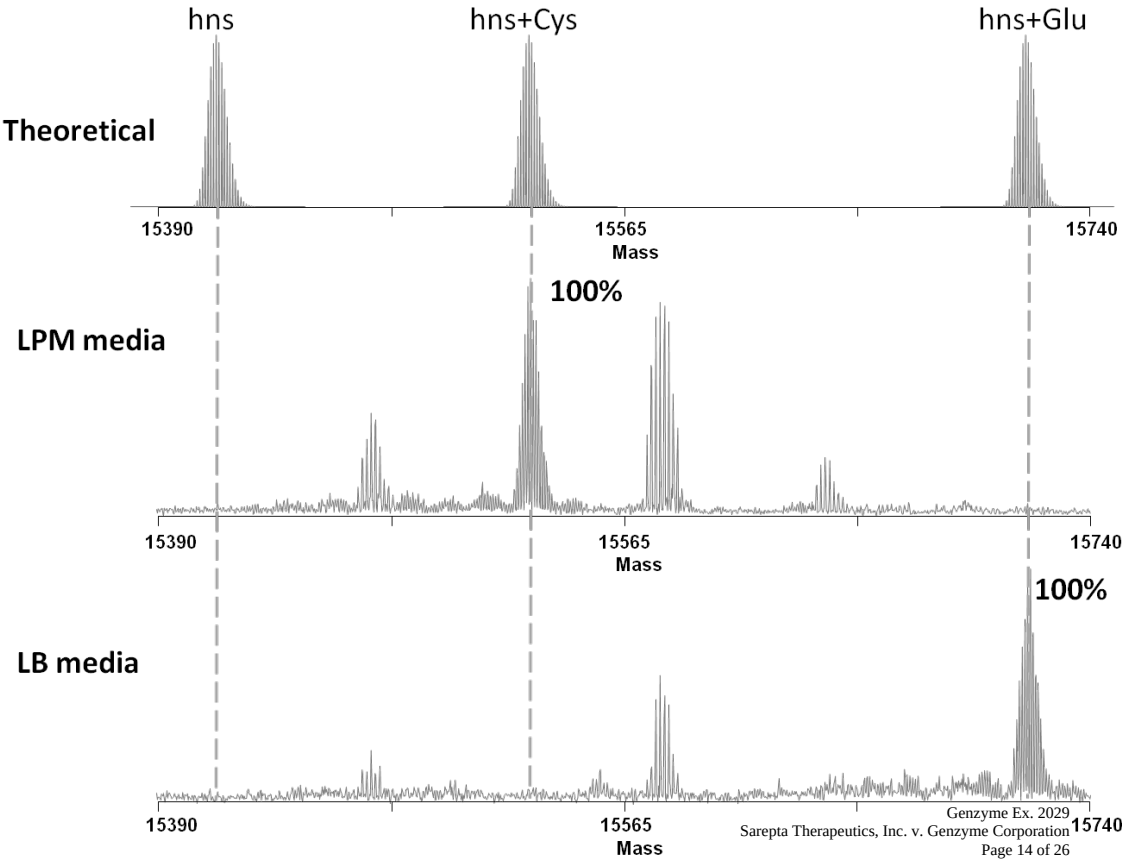
LB medium



LPM medium

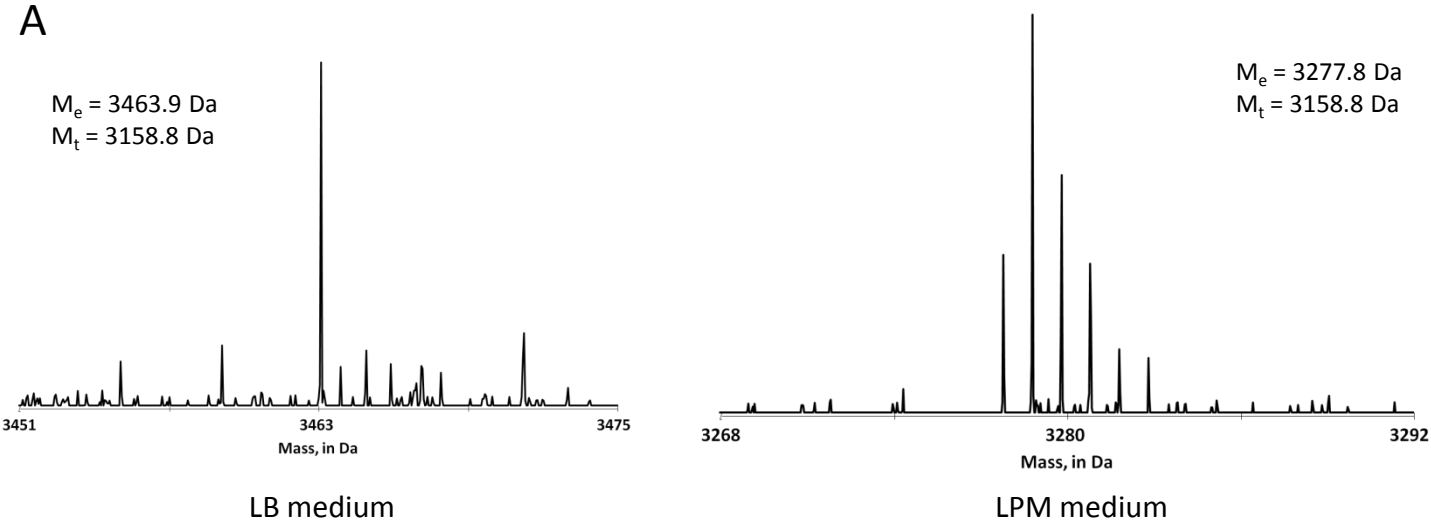


B



STM14_4143 rplC 50S ribosomal protein L3

A

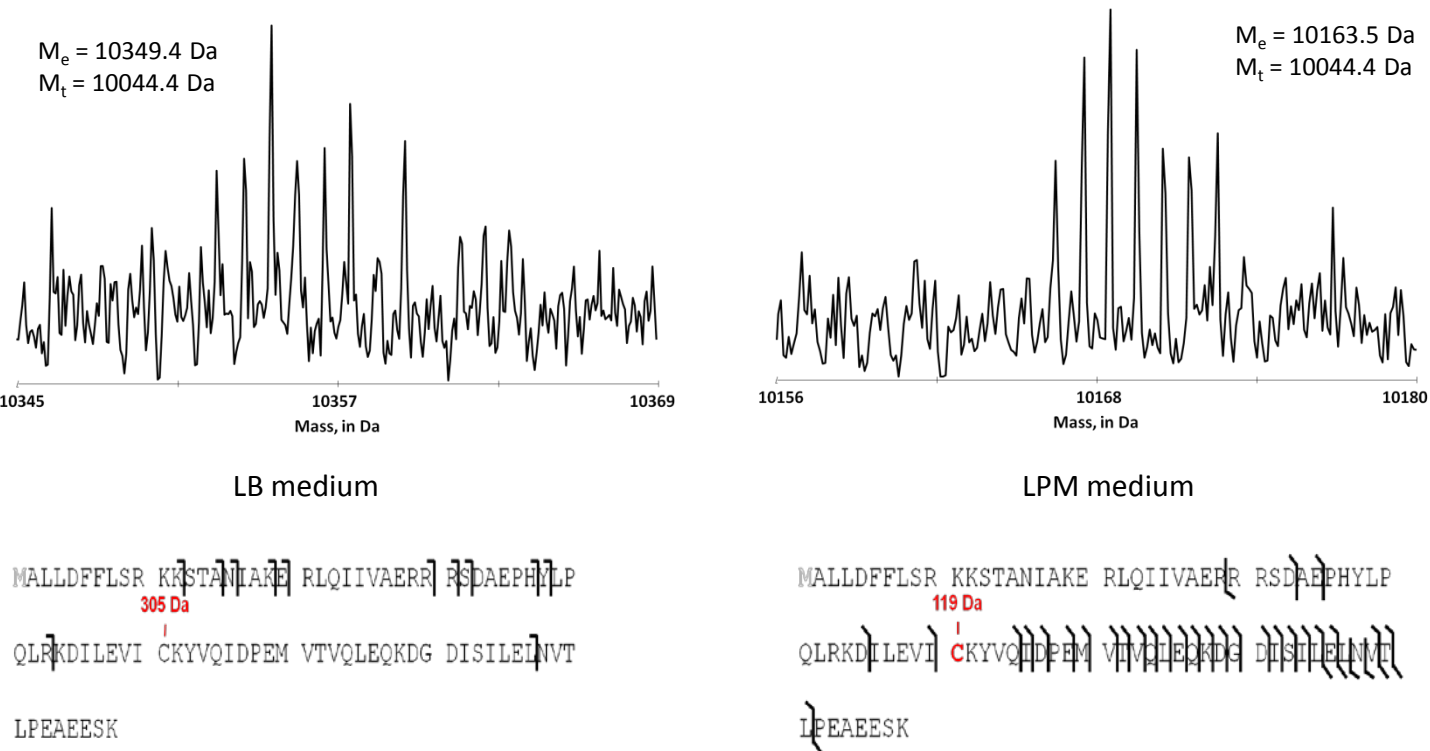


.....3 lines skipped.....
TVKRWNFRTQ DATHGNSLSH RVPGSIGQNO TPGKVFKGKK
MAGQMGNERV TVQSLDVVRV DAERNLLLVK GSVPGATGCD
LIVKPAVKA

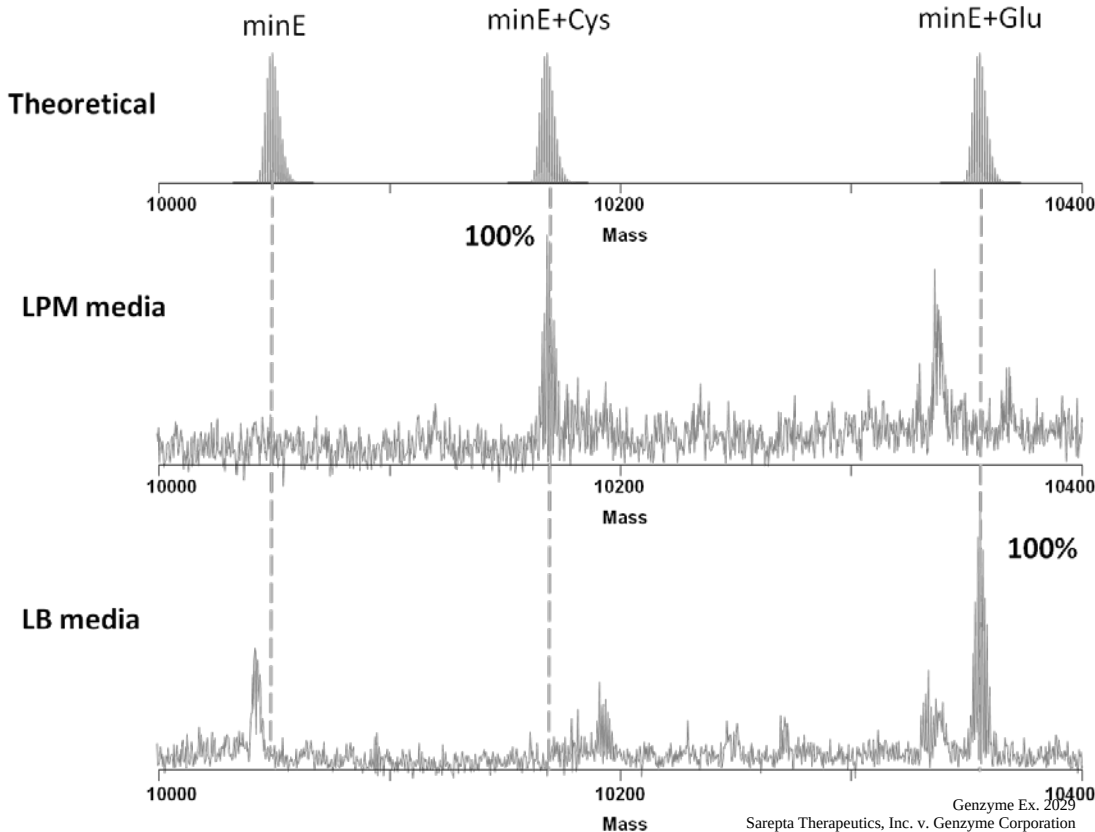
.....3 lines skipped.....
TVKRWNFRTQ DATHGNSLSH RVPGSIGQNO TPGKVFKGKK
MAGQMGNERV TVQSLDVVRV DAERNLLLVK GSVPGATGCD
LIVKPAVKA

STM14_2195 minE cell division topological specificity factor MinE

A

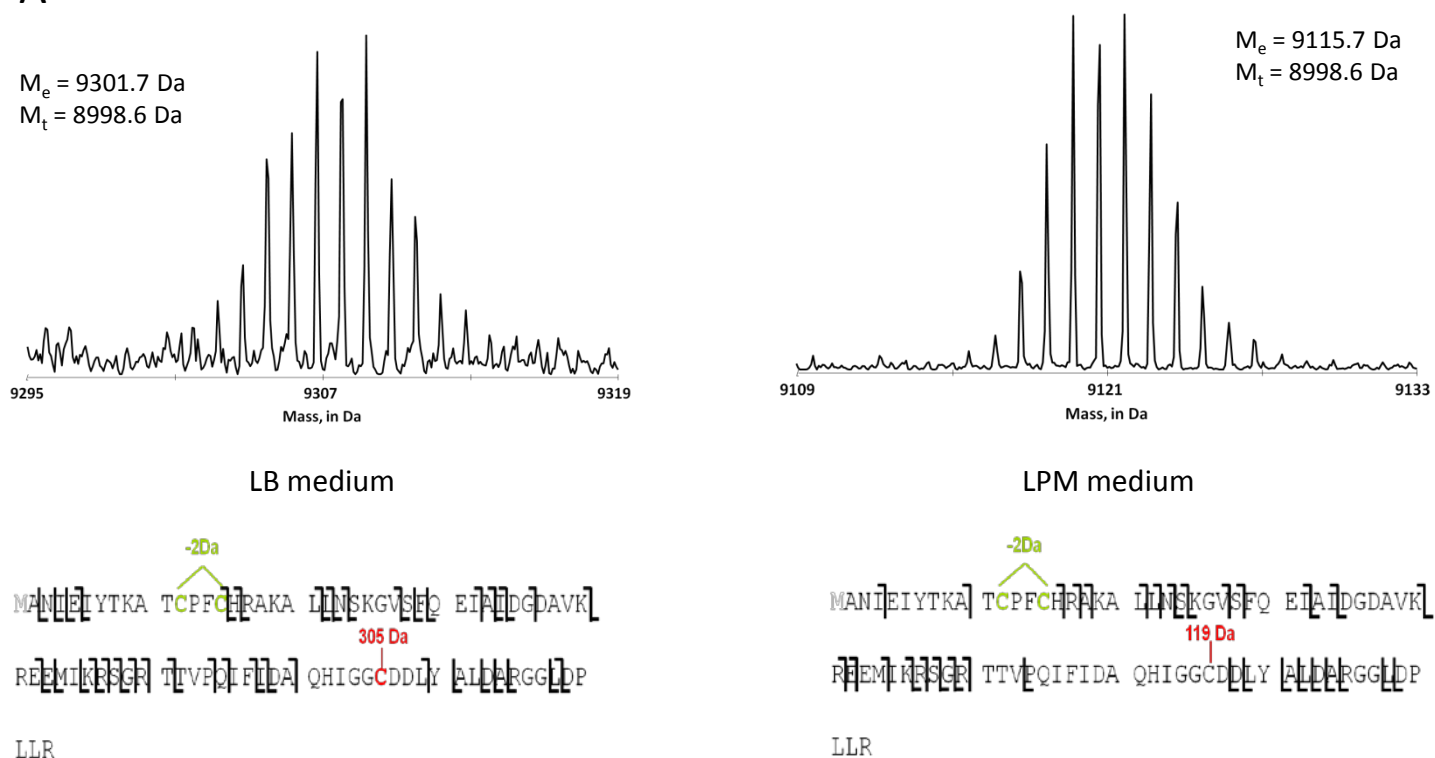


B

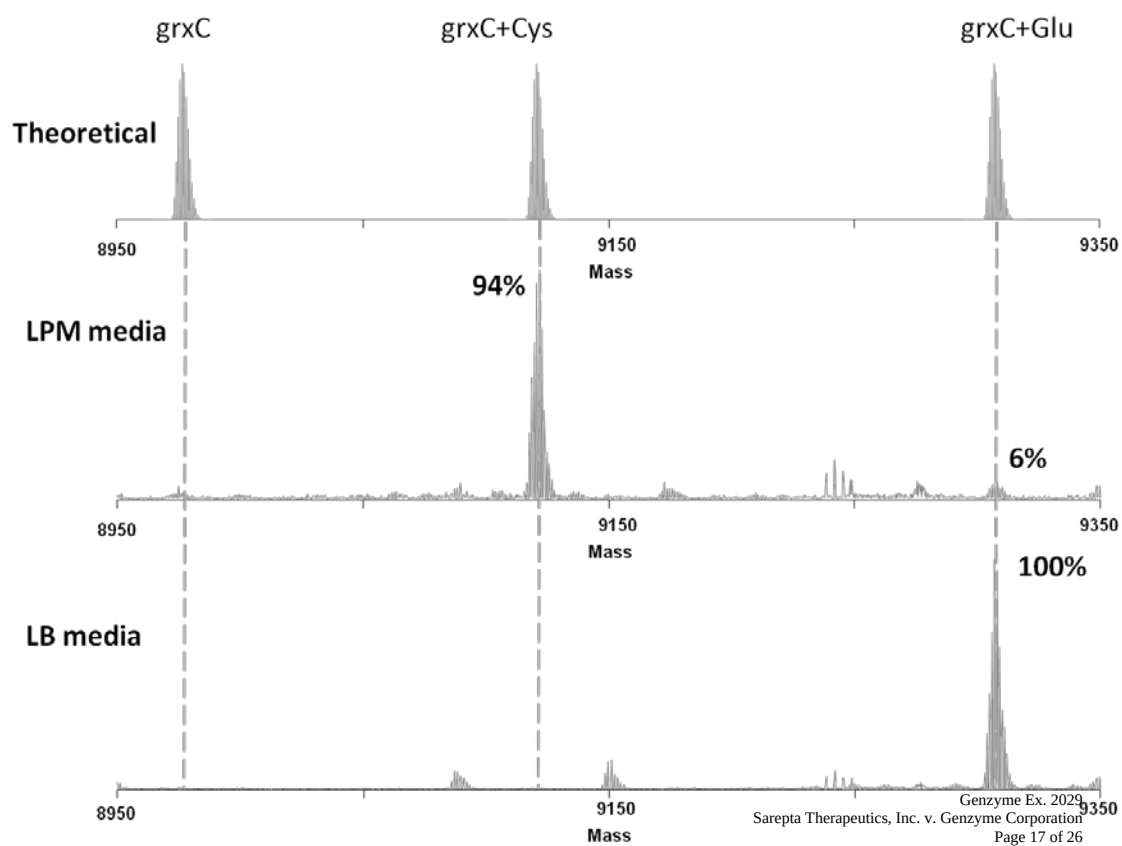


STM14_4462 grxC glutaredoxin 3

A

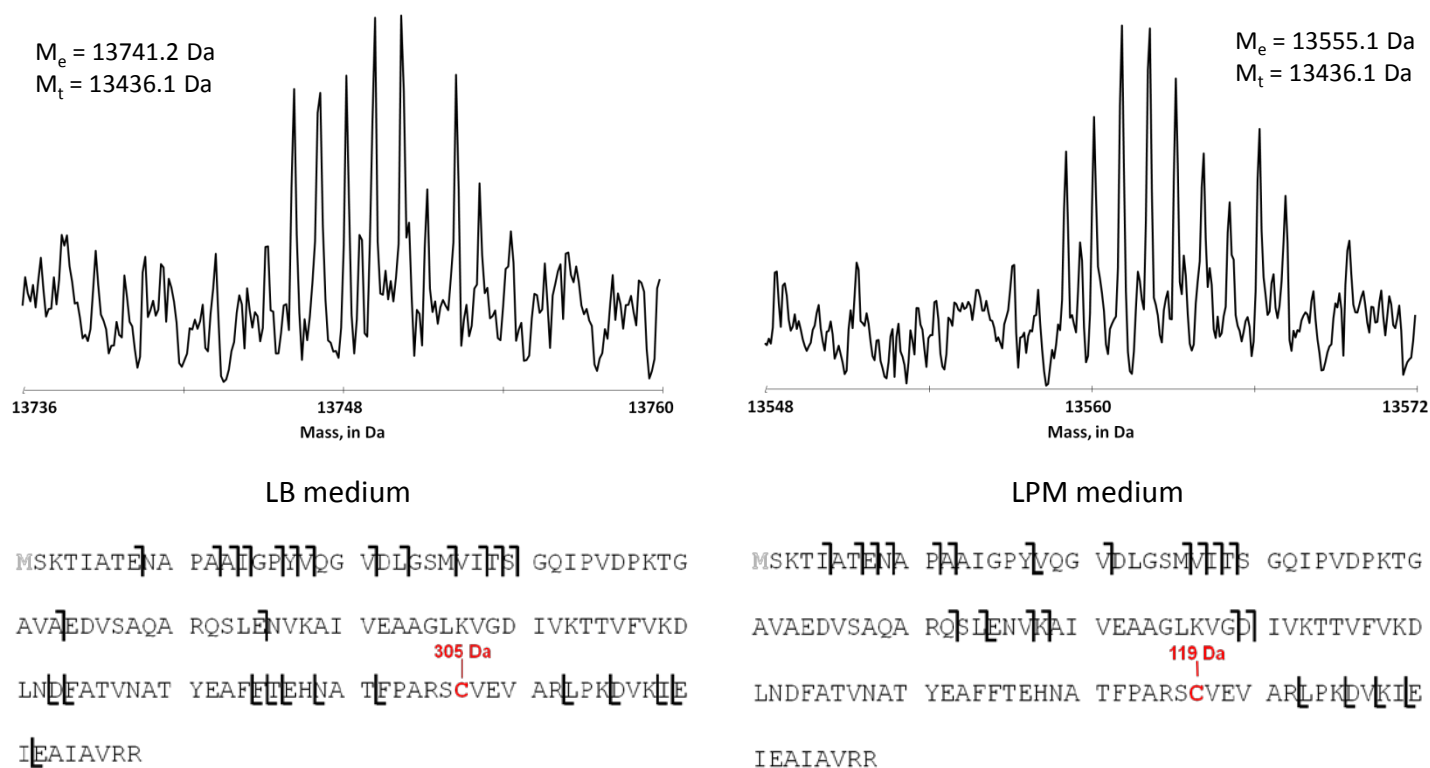


B

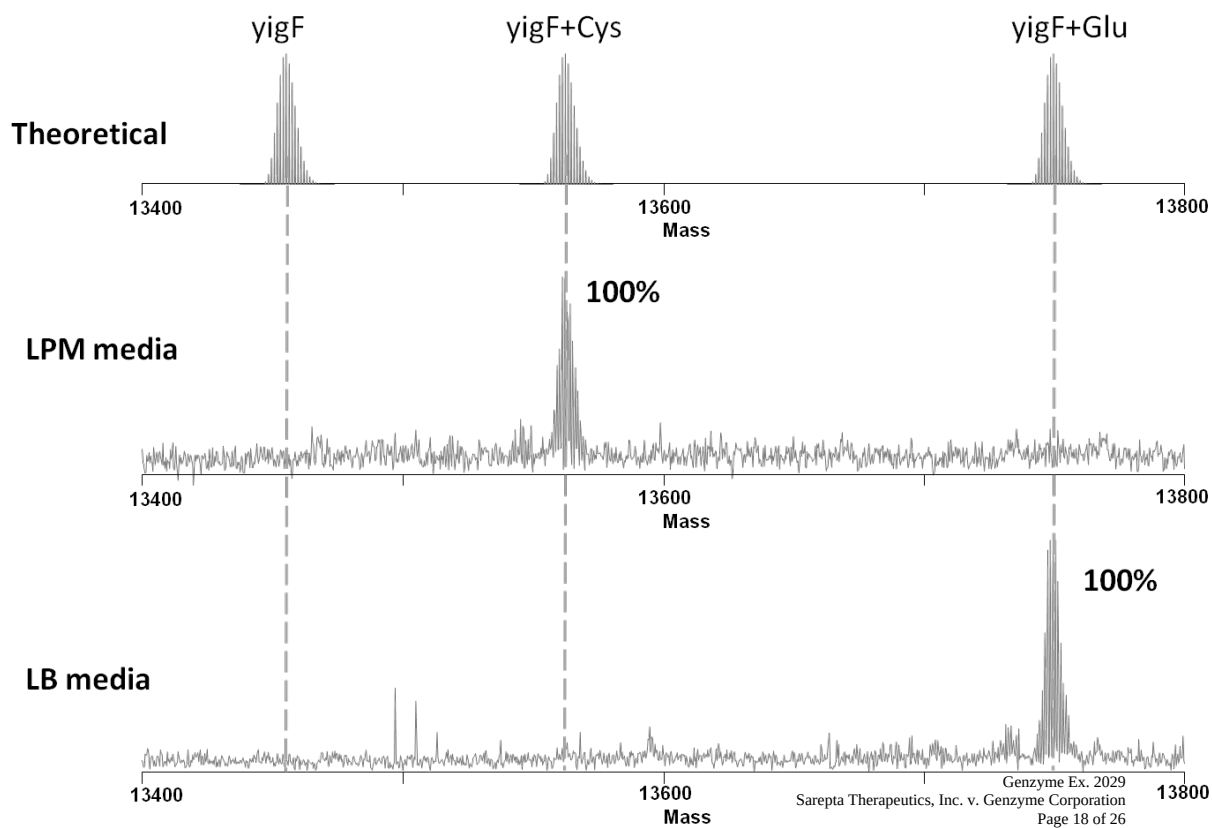


STM14_5351 yigF putative translation initiation inhibitor

A

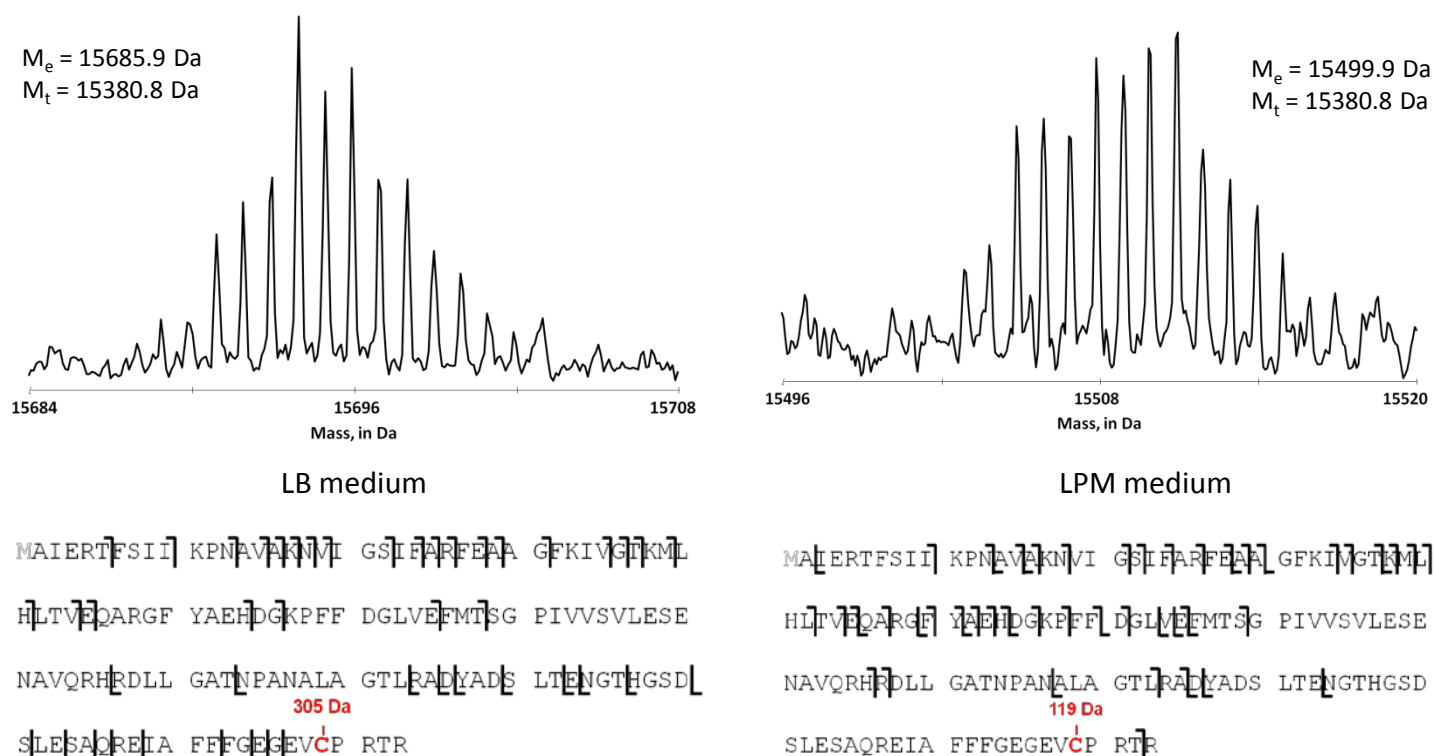


B

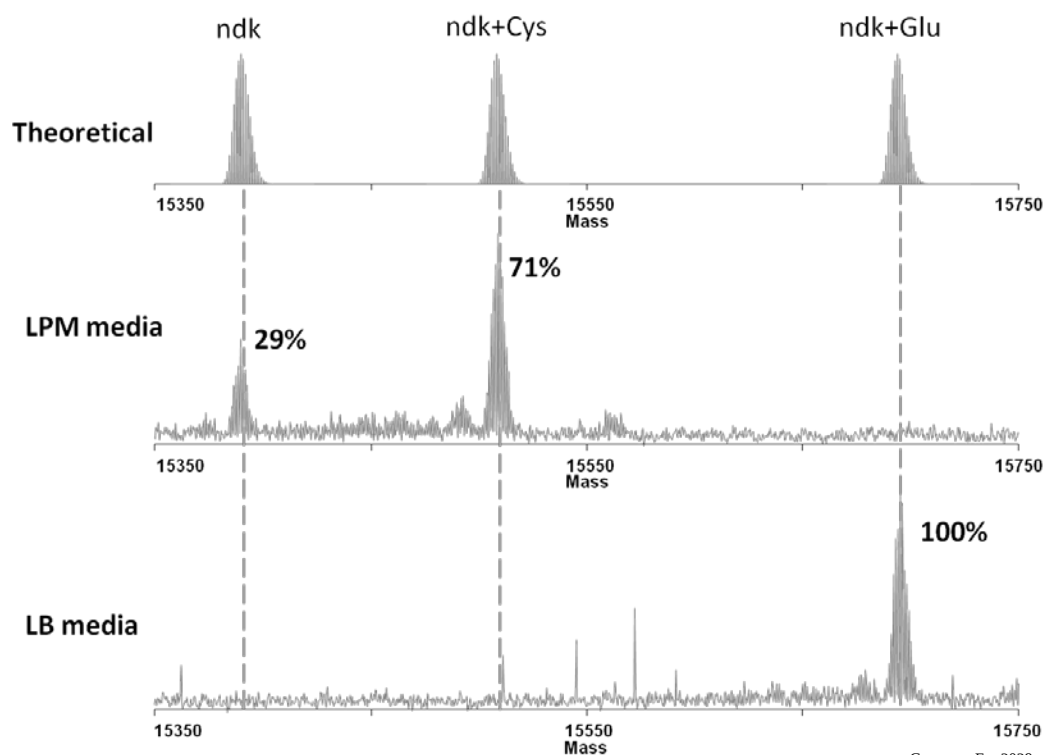


STM14_3098 ndk nucleoside diphosphate kinase

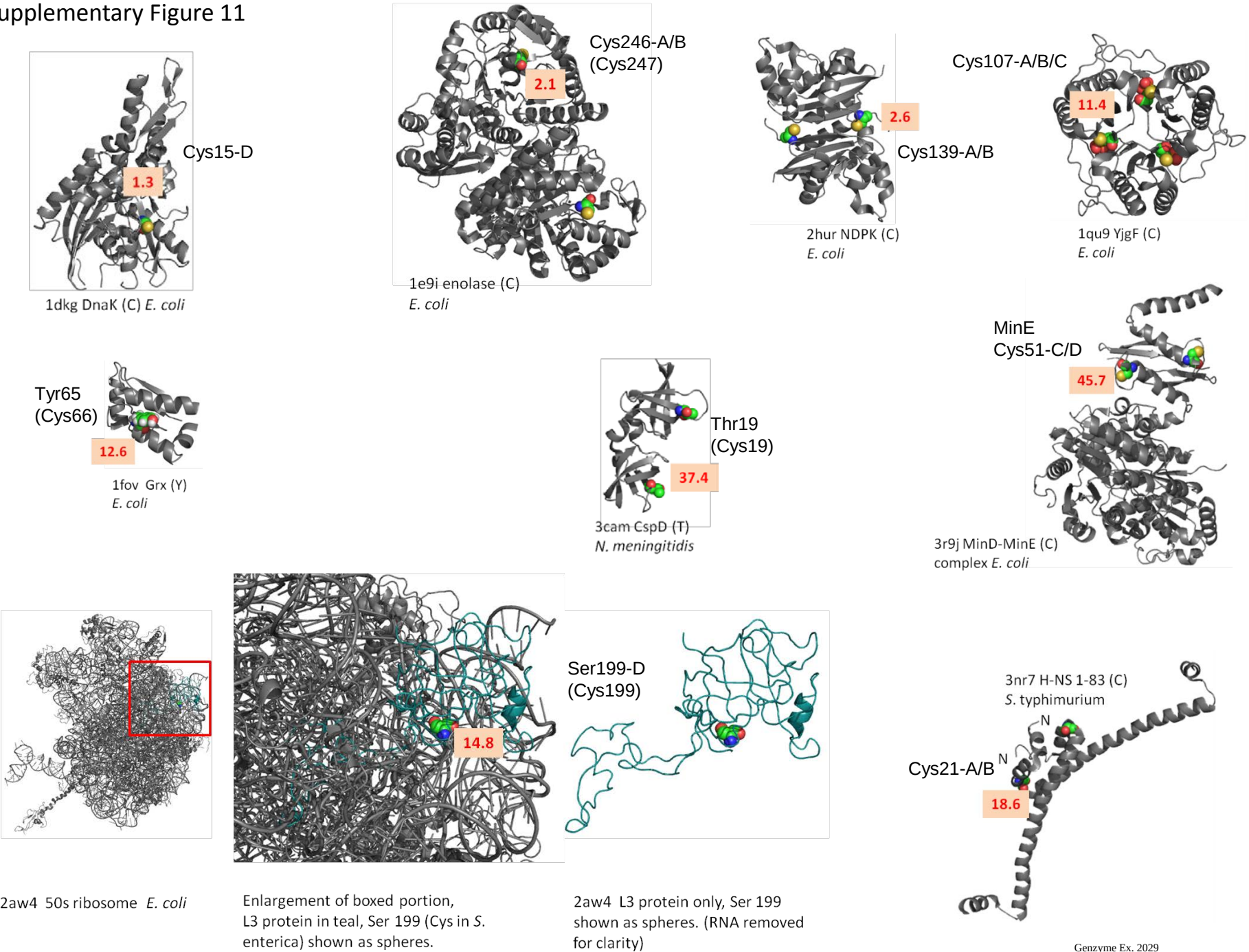
A



B



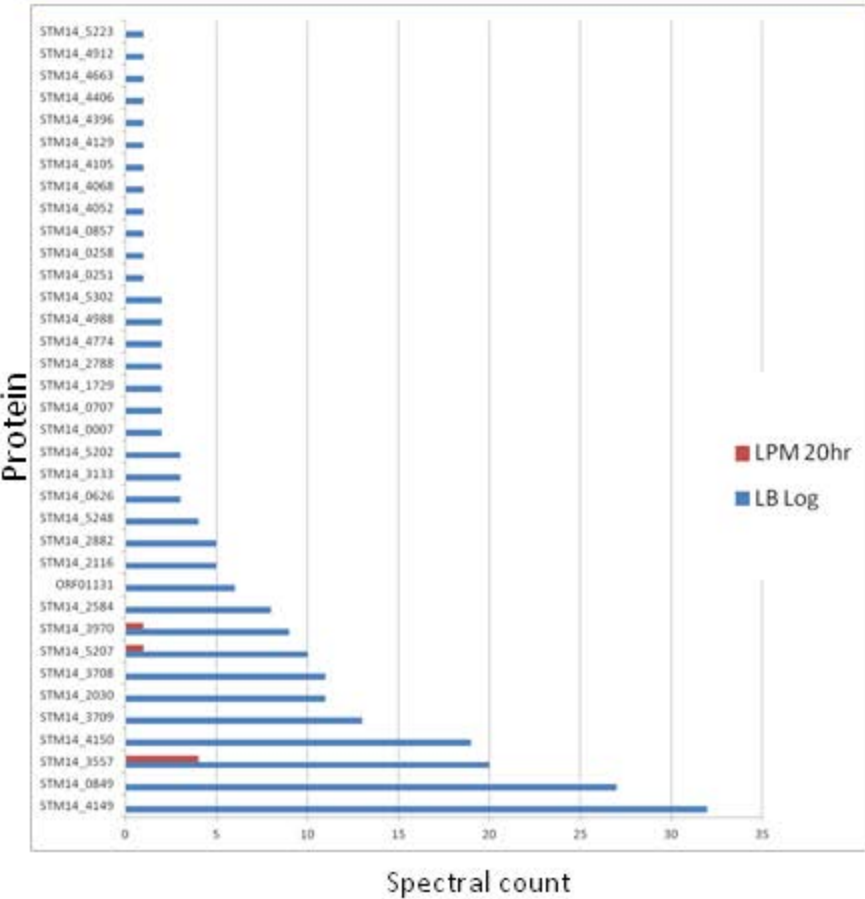
Supplementary Figure 11



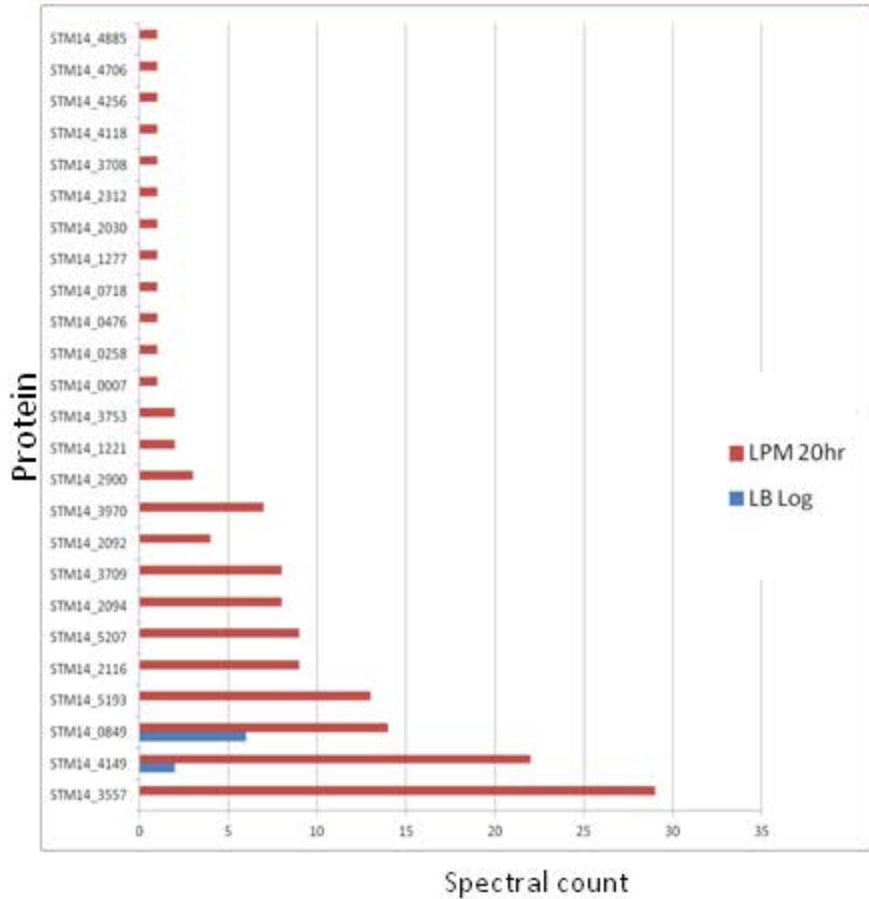


Supplementary Figure 13

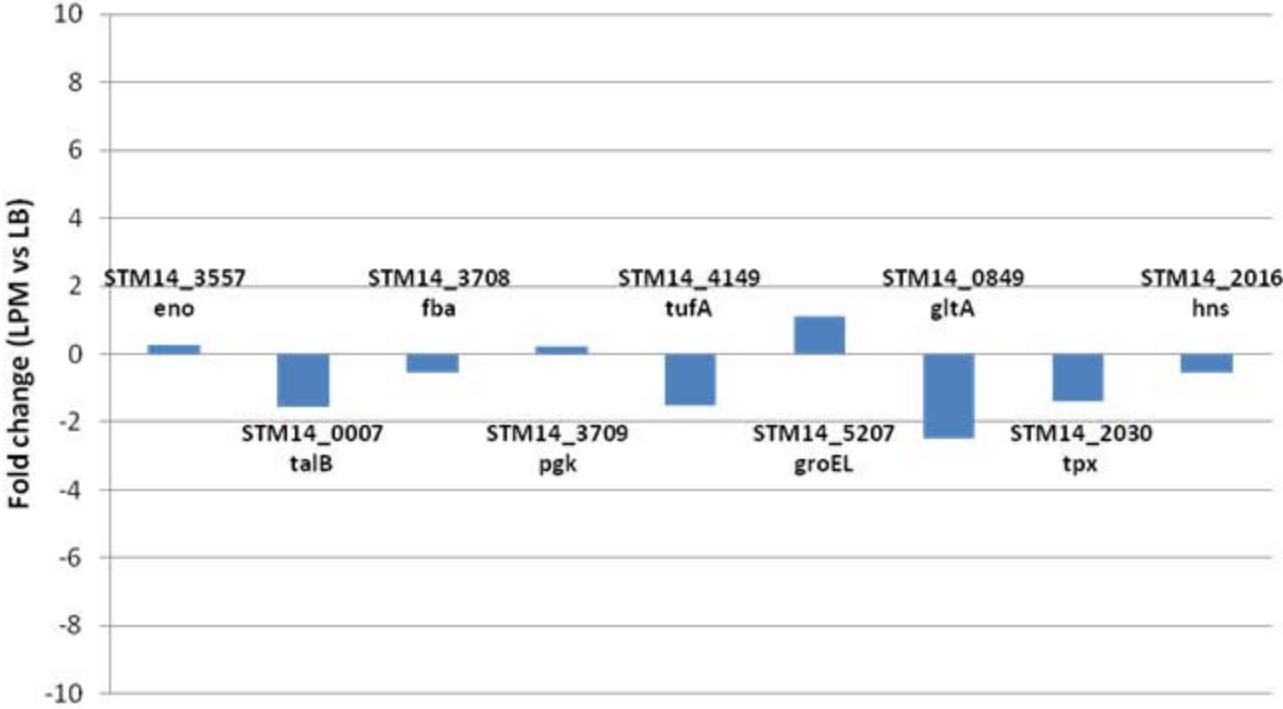
Spectral counting-based quantification of glutathionylated peptides



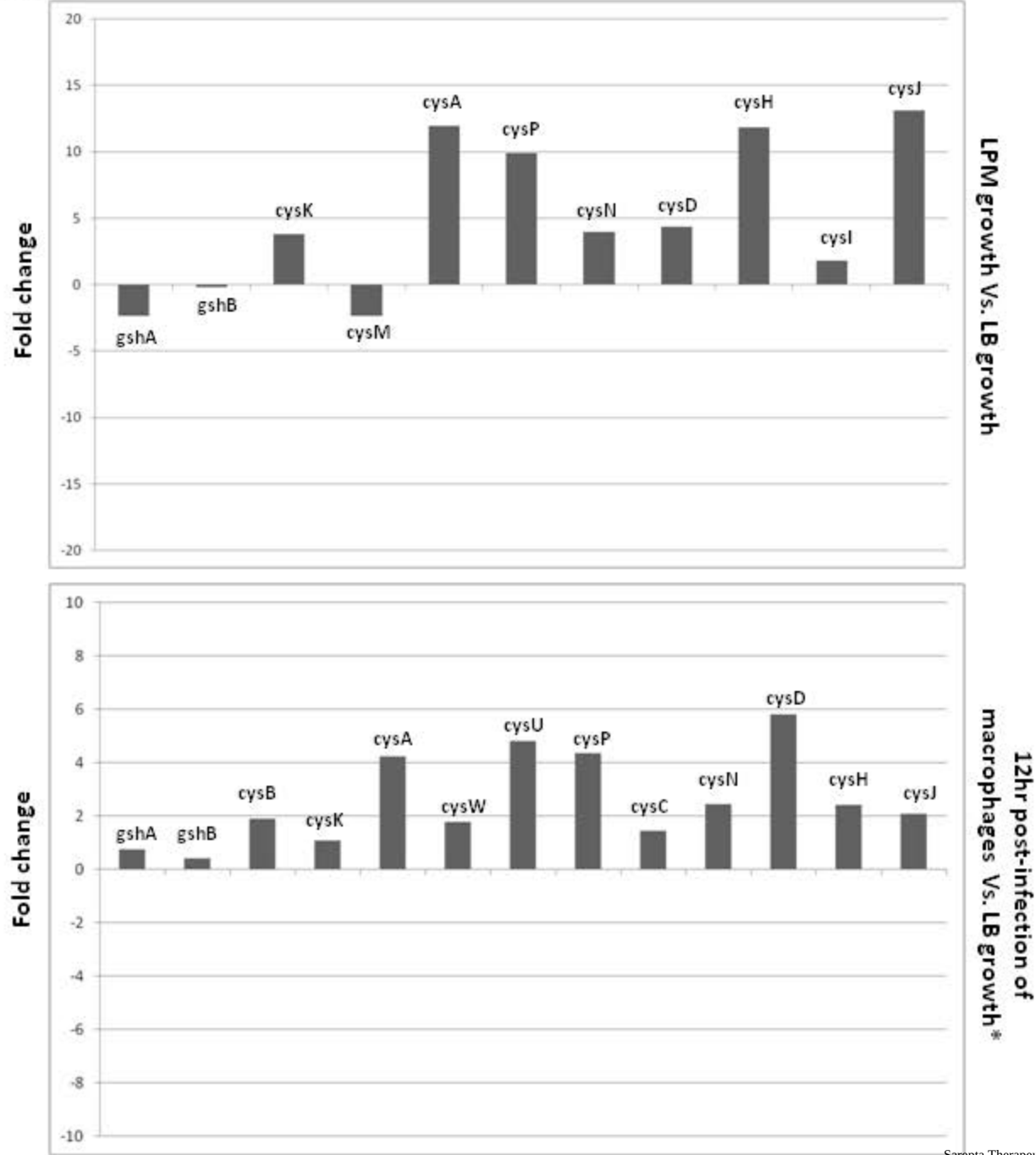
Spectral counting-based quantification of cysteinylated peptides



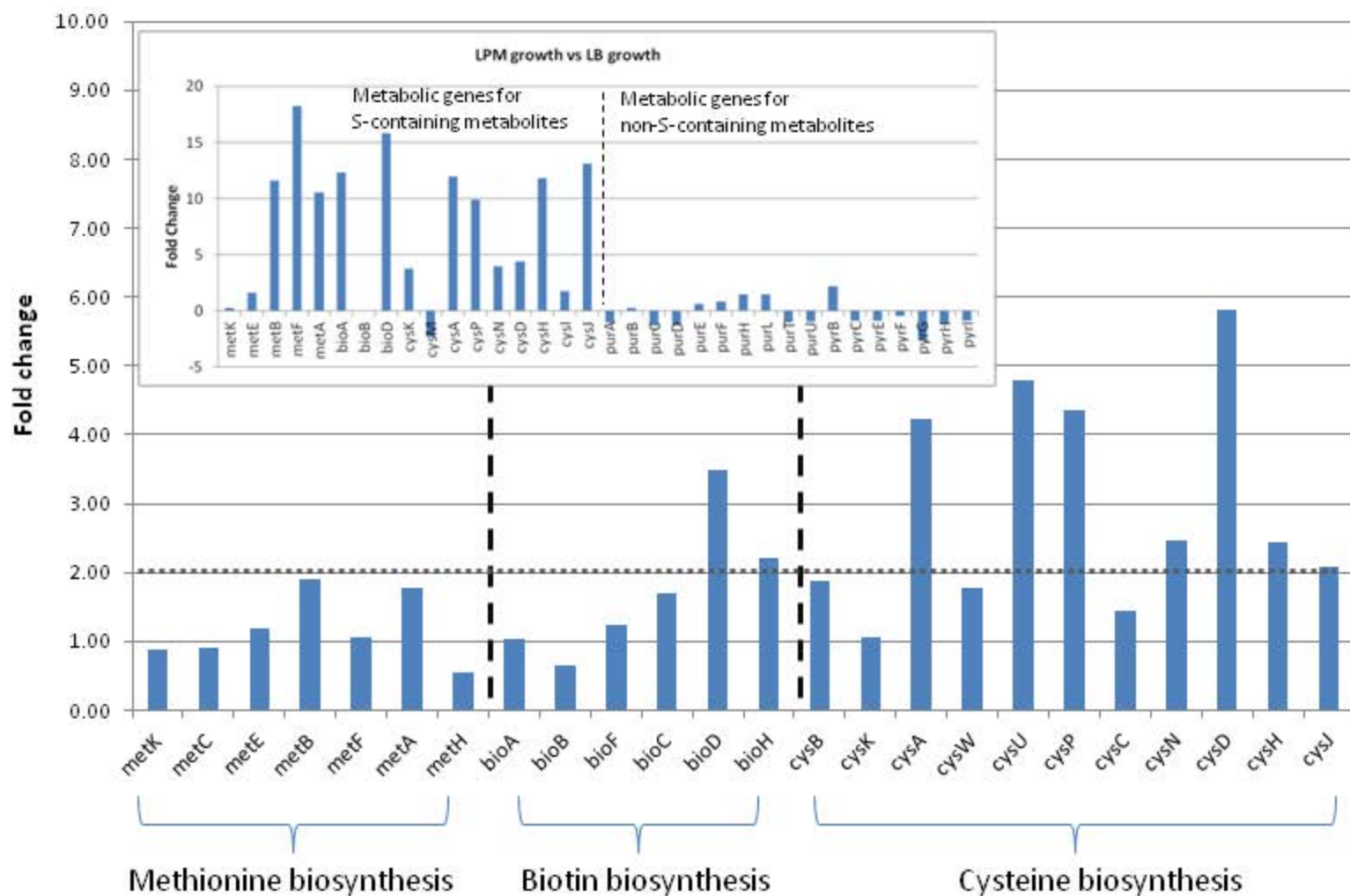
Supplementary Figure 14



Supplementary Figure 15



12hr post-infection of macrophages Vs. LB growth*



Supplementary Figure 17

