Supporting Information for:

Quantitative N-terminal Footprinting of Pathogenic Mycobacteria Reveals Differential Protein Acetylation

Cristal Reyna Thompson[1,3,4], Matthew M. Champion[2,3,4*], and Patricia A. Champion[1,3,4*]

[1]Department of Biological Sciences, [2]Department of Chemistry and Biochemistry, [3]Eck Institute of Global Health, [4]Boler-Parseghain Center for Rare and Neglected Diseases, University of Notre Dame, Notre Dame IN, USA.

Table of Contents:

Figure S-1. Thompson et al.

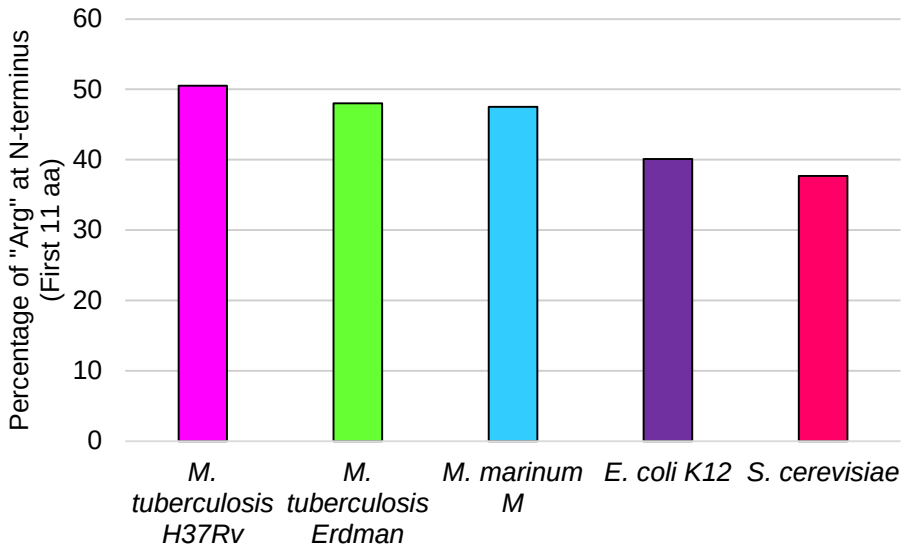
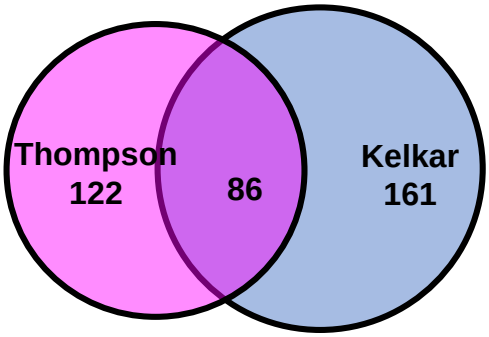


Figure S-2. Thompson et al.

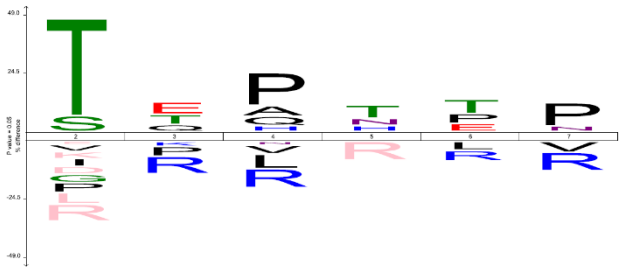
***M. tb* H37Rv**

a.



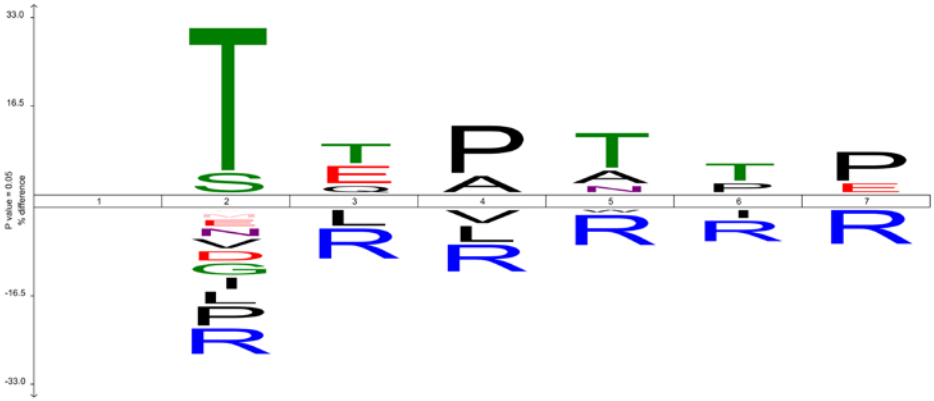
b. <u>N-terminally acetylated population</u>

Kelkar *et al*



c. <u>N-terminally acetylated population</u>

Combined data set



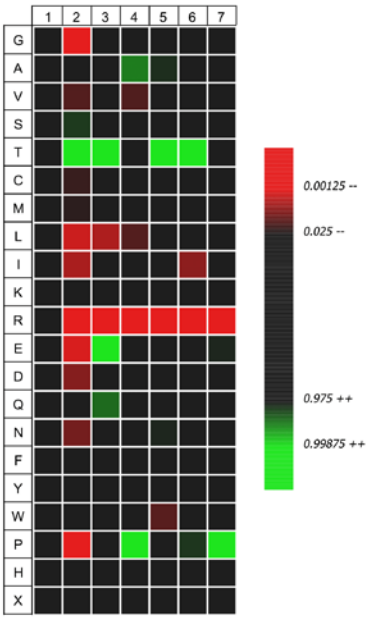


Figure S-3. Thompson et al.

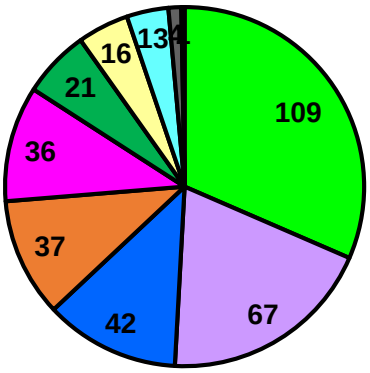
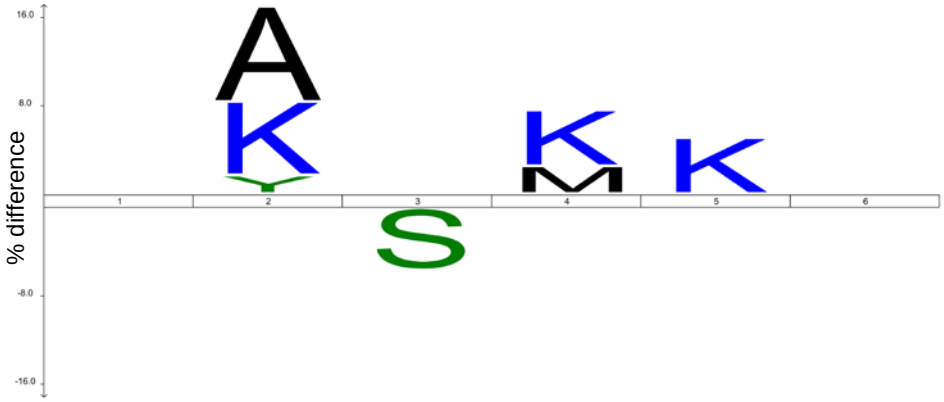
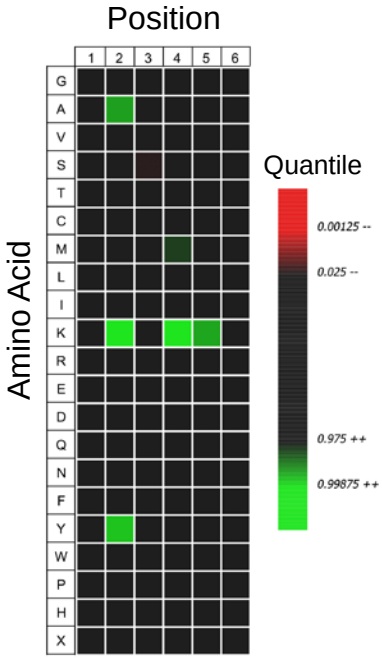


Figure S-4. Thompson et al.

a. **N-terminally acetylated peptides from *M. marinum***

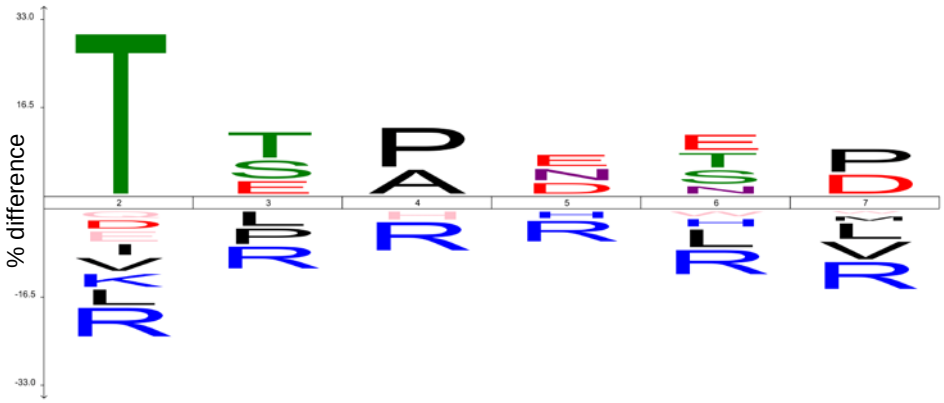
Met in position 1





b.

iMet cleaved



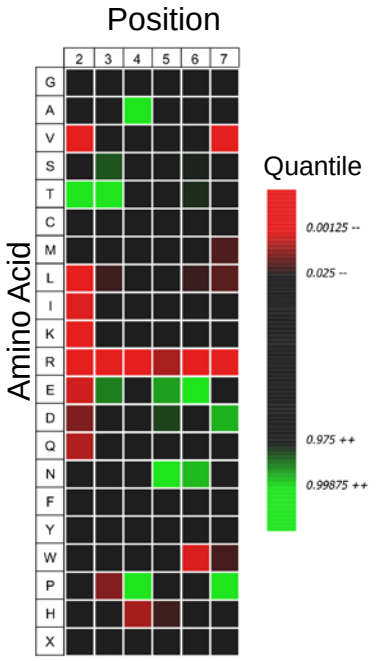
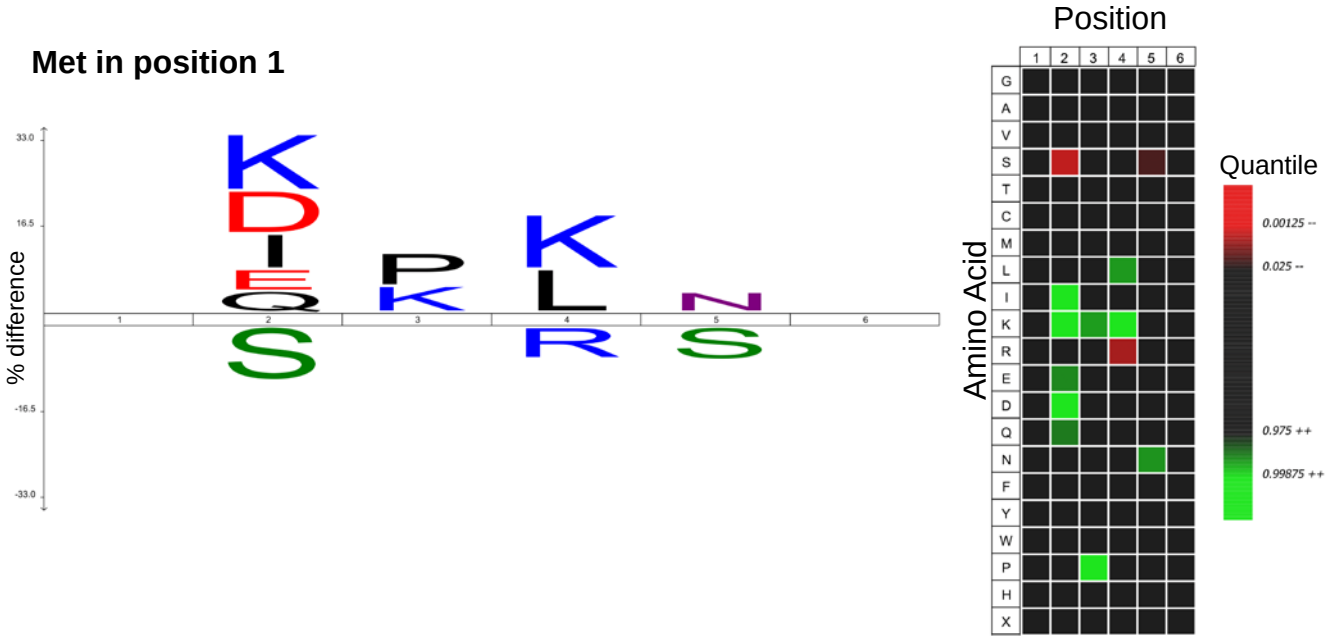


Figure S-5. Thompson et al.

a. **Heavy N-terminally acetylated populations from *M. tb***

Met in position 1



b.

iMet cleaved

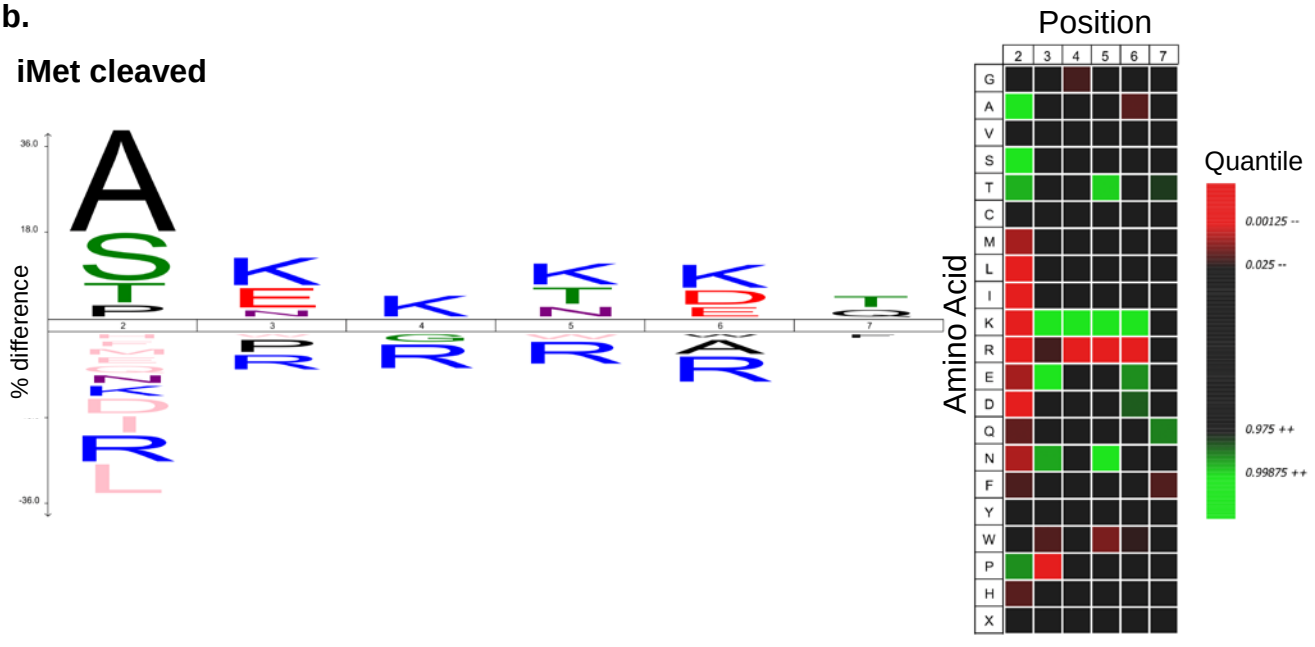
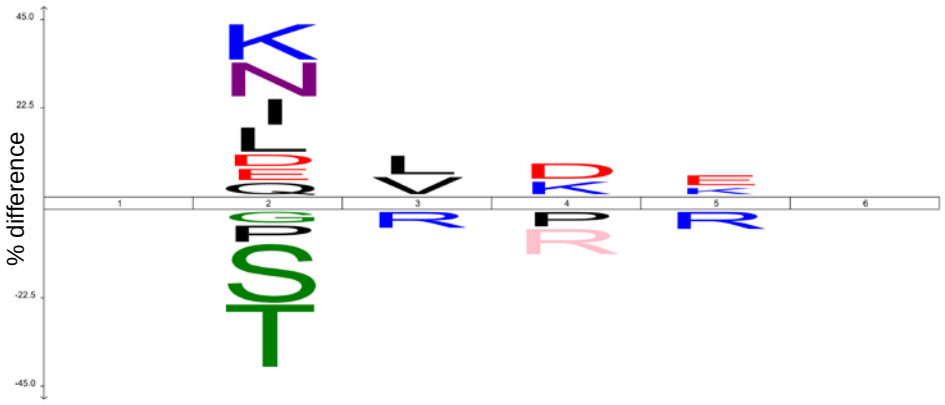
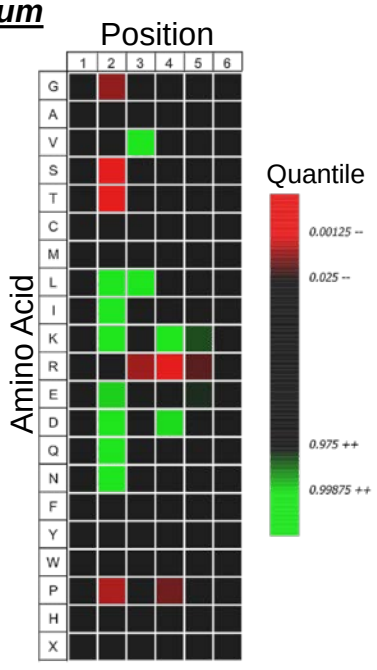


Figure S-6. Thompson et al.

a. **Heavy N-terminally acetylated populations from *M. marinum***

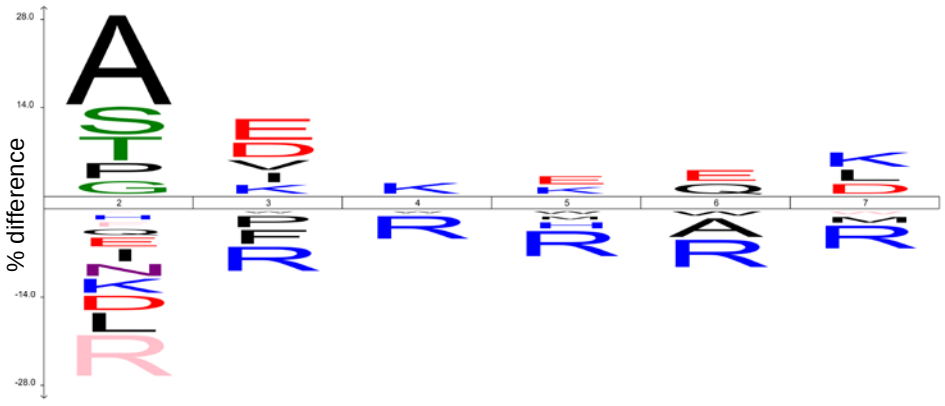
Met at position 1





b.

iMet cleaved



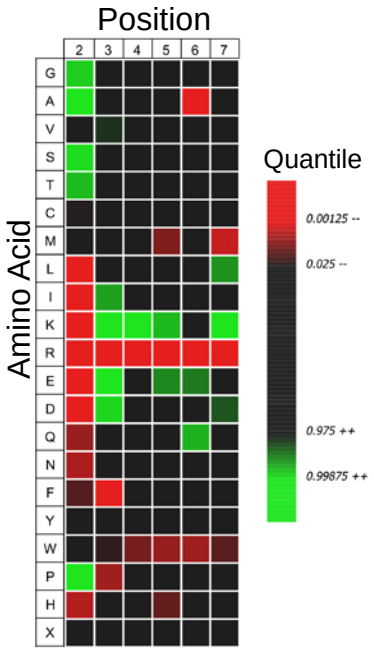
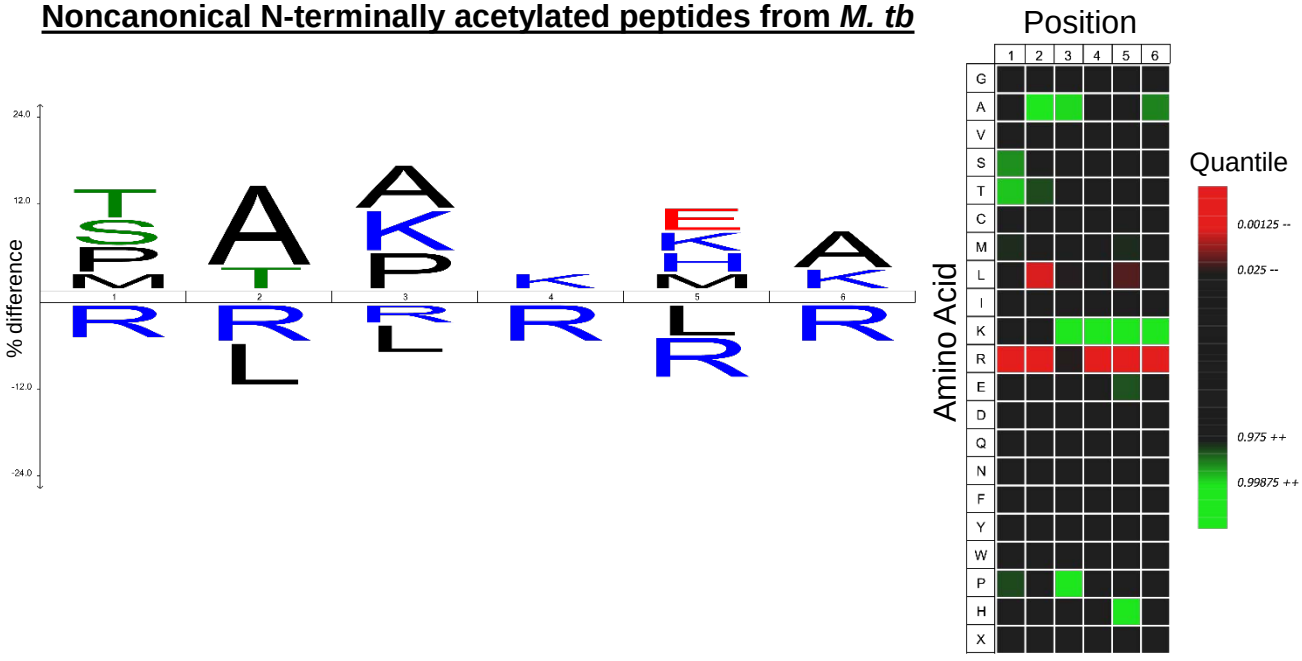


Figure S-7. Thompson et al.



Supplemental Figure Legends

Figure S-1. Arg frequency in canonical proteomes. Proteins from each organism were analyzed for Arg frequency in the first 11 amino acids of the canonical amino acid of each protein. Frequencies are reported as the number of proteins containing one Arg or more in the first 11 amino acids out of the total number of proteins for the respective organism.

Figure S-2. Comparison and analysis of *M. tb* H37Rv N-terminally acetylated proteins identified within the Kelkar et al data set (2011). a. Comparison of N-terminally acetylated protein IDs generated in this study as compared to the Kelkar et al. study [1]. **b**. Average N-terminal motif of N-terminally acetylated proteins at P2' in the Kelkar study **c**. Visualization of the average N-terminal motif of combined data sets using the N-terminal analysis method in IceLogo [1]. Heatmaps (right) for each population are representative of significant amino acids ($p < 0.05$). Green denotes enrichment and red denotes underrepresentation of amino acids at the designated polypeptide position.

Figure S-3. Functional characterization of *M. marinum* N-terminally acetylated proteins. a. Mycobrowser was used to identify the functional categories of the N-terminally acetylated proteins identified in *M. marinum*.

Figure S-4. IceLogo analysis of N-terminally acetylated protein populations for *M. marinum* reveal distinct N-terminal motifs for acetylation. Peptides identified as N-terminally acetylated were analyzed against the canonical *M. marinum* proteome by IceLogo [1]. **a.** 88 N-terminally acetylated peptides with intact iMet and **b.** 272 N-terminally acetylated peptides with cleaved iMet were analyzed by IceLogo using the N-terminal analysis method. Heatmaps (right) for each population are representative of significant amino acids ($p < 0.05$). Green denotes enrichment and red denotes underrepresentation of amino acids at the designated polypeptide position.

Figure S-5. IceLogo analysis of non-acetylated protein populations for *M. tb.* Peptides recovered with exogenous acetylation at the protein N-terminus were analyzed against the canonical proteome of *M. tb* by IceLogo [1]. **a.** 79 exogenously acetylated peptides with intact iMet and **b.** 303 exogenously acetylated peptides with cleaved iMet were analyzed by IceLogo using the N-terminal analysis method. Heatmaps (right) for each population are representative of significant amino acids ($p < 0.05$). Green denotes enrichment and red denotes underrepresentation of amino acids at the designated polypeptide position.

Figure S-6. IceLogo analysis of non-acetylated protein populations for *M. marinum.* Peptides recovered with exogenous acetylation at the protein N-termini were analyzed against canonical protein N-terminus of the *M. marinum* proteome. **a.** 163 exogenously acetylated peptides with intact iMet and **b.** 537 exogenously acetylated peptides with cleaved iMet were analyzed by IceLogo [1] using the N-terminal analysis method. Heatmaps (right) for each population are representative of significant amino acids ($p < 0.05$). Green denotes enrichment and red denotes underrepresentation of amino acids at the designated polypeptide position.

Figure S-7. IceLogo analysis of non-canonical acetylated protein populations for *M. tb.* Peptides identified with N-terminal acetylation at non-canonical N-termini were screened against trypsin and GluC N-terminal cleavages. 186 peptides were analyzed by IceLogo using the random analysis method. Heatmap (right) are representative of significant enrichment or depletion of amino acids ($p < 0.05$). Green denotes enrichment and red denotes underrepresentation of amino acids at designated polypeptide position.

Supplemental Methods

For the Supplemental Figures 2-5, BioCyc, a collection of pathway and genome databases, was used to identify the H37Rv orthologs for *M. tb* Erdman (biocyc.org). Mycobrowser [[2], http://mycobrowser.epfl.ch/marinolist.html] was used to identify the H37Rv orthologs for *M. marinum.* Mycobrowser was also used to assign functional categories for N-terminally acetylated proteins of *M. tuberculosis* H37Rv and *M. marinum* M. The functional categories are fully described in Cole et al[3].

Supplemental References

[1] Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo, *Nat Methods 6*, 786-787.

[2] Kapopoulou, A., Lew, J. M., and Cole, S. T. (2011) The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes, *Tuberculosis (Edinb) 91*, 8-13.

[3] Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S., and Barrell, B. G. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence, *Nature 393*, 537-544.