Systems proteomics approaches to study bacterial pathogens: application to Mycobacterium tuberculosis

Review Article

Author(s):
Banaei-Esfahani, Amir; Nicod, Charlotte; Aebersold, Ruedi; Collins, Ben C.

Publication date:
2017-10

Permanent link:
https://doi.org/10.3929/ethz-b-000201963

Originally published in:
Current Opinion in Microbiology 39, https://doi.org/10.1016/j.mib.2017.09.013

Funding acknowledgement:
161435 - The host-pathogen protein interactome of Mycobacterium tuberculosis (SNF)

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Systems proteomics approaches to study bacterial pathogens: application to *Mycobacterium tuberculosis*

Amir Banaei-Esfahani[1,2], Charlotte Nicod[1,2], Ruedi Aebersold[1,3,4] and Ben C Collins[1,4]

Significant developments and improvements in basic and clinical research notwithstanding, infectious diseases still claim at least 13 million lives annually. Classical research approaches have deciphered many molecular mechanisms underlying infection. Today it is increasingly recognized that multiple molecular mechanisms cooperate to constitute a complex system that is used by a given pathogen to interfere with the biochemical processes of the host. Therefore, systems-level approaches now complement the standard molecular biology techniques to investigate pathogens and their interactions with the human host. Here we review omic studies in *Mycobacterium tuberculosis*, the causative agent of tuberculosis, with a particular focus on proteomic methods and their application to the bacilli. Likewise, the discussed methods are directly portable to other bacterial pathogens.

Addresses
[1] Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland
[2] PhD Program in Systems Biology, Life Science Zurich Graduate School, University of Zurich and ETH Zurich, Zurich, Switzerland
[3] Faculty of Science, University of Zurich, Zurich, Switzerland

Corresponding authors: Aebersold, Ruedi (aebersold@imsb.biol.ethz.ch), Collins, Ben C (collins@imsb.biol.ethz.ch)
[4] These authors jointly supervised this work.

Current Opinion in Microbiology 2017, **39**:64–72

This review comes from a themed issue on **Bacterial systems biology**

Edited by **Christoph Dehio** and **Dirk Bumann**

For a complete overview see the Issue and the Editorial

Available online 13th October 2017

http://dx.doi.org/10.1016/j.mib.2017.09.013

Introduction

Thirty six years ago, improvement in hygiene and nutrition as well as the development of new drugs and vaccines conferred a dominant thought that infectious diseases are close to becoming insignificant [1]. However, due to the persistence of serious infectious diseases the death rate attributed to infection now stands at 13–15 million people per year and will continue as such at least until 2030 [2]. *Mycobacterium tuberculosis* (Mtb) is the leading cause of mortality among infectious diseases and claims ~1.8 million lives annually. Macrophages recognize and engulf the bacilli following infection but cannot thoroughly clear them, making Mtb difficult to treat. Mtb achieve this intracellular state by modulating the immune and other cellular systems at various levels. For instance, Mtb partially blocks phagosomal acidification, a prerequisite for phagosomal rupture, and lysosome–phagosome fusion following uptake by macrophages [3]. A growing body of evidence suggests that the strategy applied by a pathogen such as Mtb to subvert the human host involves multiple mechanisms and that systems-level studies are indispensable to study them. In the Mtb field, the vast majority of effort in systems-level analyses have been focused on genomic and functional genomics studies. Yet, proteins are the main functional elements of biochemical pathways, thus determining phenotypic traits. The state of the proteome of a cell is therefore expected to provide more direct functional information than genomic analyses alone. Over the past decade, new proteomic methods and tools have been developed that now support the quantification of proteins systematically and reproducibly over many samples, thus facilitating systems-level studies at the level of the proteome.

Lessons from genetic and transcriptional studies of Mtb

Mtb, as an intracellular pathogen has a GC rich genome of 4.4 million base pairs that contains 4018 protein coding genes. The genome was initially sequenced in 1998 [4]. Strikingly, 26% of the Mtb genes, called leaderless genes, lack 5′UTR and hence the Shine-Dalgarno sequence commonly used for the initialization of ribosome engagement [5]. The general lack of genetic recombination in bacteria and horizontal gene transfer keep the Mtb genome in complete linkage [6,7]. Advances in DNA sequencing throughput and affordability paved the way to explore the diversity and evolution of the bacilli with respect to the human host. Genomic results from clinical cohort studies suggest that Mtb has co-evolved, migrated and expanded within the human host, since its origin in Africa [8]. To date, Mtb's phylogenetic dendrogram comprises seven lineages of strains infecting humans. They are associated with different geographical regions, with any two strains differing by ~1200 SNPs on average [9]. Contrary to the situation in many other bacteria, about two thirds of the SNPs of Mtb in the coding regions are non-synonymous and more than half fall into highly conserved positions, hinting that the majority of the

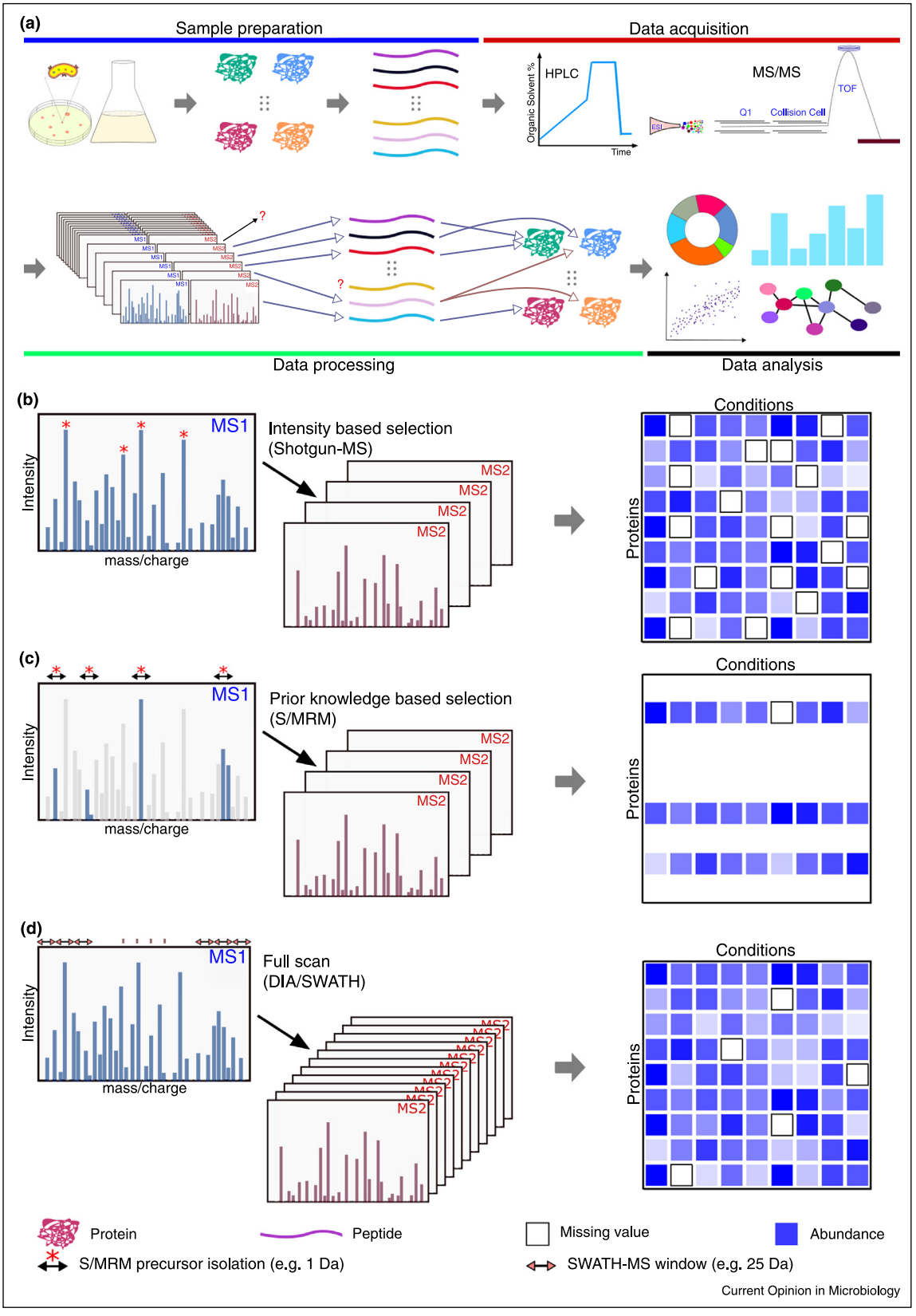
mutations are functional and consequently implicated in the phenotypic diversity of the pathogen [10,11]. Clinical isolates of Mtb feature diverse phenotypes in response to the host [7,12••]. Despite the fact that pathogens usually carry variable antigens to evade the host immune system, human T-cell epitopes in Mtb are highly conserved, suggesting that lung damage during tuberculosis might be largely caused by the host immune response which increases patient coughing and eventually facilitates the transmission of the pathogen. Genome-wide association studies (GWAS) together with standard molecular biology techniques identified the most common causative mutations conveying resistance to at least 16 different drugs [13,14,15•,16•]. Also, recently, a genomic study on 5310 Mtb genomes elucidated that the katG mutation encoding p.S315T, which confers resistance to isoniazid, overwhelmingly arises before causative mutations of rifampicin resistance across all the lineages. In spite of these and other successes of genomic studies, it remains challenging to derive detailed molecular mechanisms of infection and persistence from such data alone.

In addition to large-scale genome sequencing efforts Mtb researchers have turned to functional genomic approaches, primarily mRNA analysis by microarrays or RNA sequencing following perturbation experiments to increase systems level understanding. For example, Mtb can effectively respond to stresses such as hypoxia and remain in a non-replicating state called dormancy, even for decades. The dormant state of Mtb introduces a marked phenotypic drug resistance and its persistence within the host. Sherman and colleagues used transcriptome analyses to demonstrate that a transcription factor, DosR (Rv3133c), regulates ~50 genes in the so-called DosR regulon in response to hypoxia [17,18]. This first effort was followed by a larger study where they mapped the transcriptional network of Mtb using ChIPSeq combined with expression data from the induction of the same transcription factors [19,20]. The data revealed Rv0081 as one of the largest hubs orchestrating the transcriptional network in hypoxic stress [21]. A related study showed the significance of two transcription factors, Rv0324 and Rv0880, in pushing Mtb into a tolerant state following the treatment with the antitubercular drug bedaquiline [22•]. However, some regulatory mechanisms are not apparent from transcriptional results. For instance, the vaccine strain of Mtb, BCG, modifies 40 ribonucleosides in tRNA in response to hypoxia which results in the selective translation of mRNAs from families of codon-biased persistence genes [23••]. Although Mtb exposed to nitric oxide stress respond rapidly at the transcriptome level, it takes some time of these changes to be revealed on protein level, a behavior that has been linked to protein degradation rather than proteins synthesis [24]. It has therefore been suggested that proteomic data should be integrated with results from other large-scale biomolecular studies and specific functional assays to generate mechanistic models of complex processes. As a way of multi-omics data integration, genetic association studies are portable to the transcriptome, metabolome and proteome of a given organism to determine how genomic variants translate into altered quantitative biomolecular profiles that eventually determine phenotypic traits [25,26,27••,28].

Proteomics methods to provide mechanistic insights in bacterial pathogens

Proteomics aims at characterizing the state of the proteome across conditions and biological samples at a given time point. It offers a wide variety of methods relying primarily on liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS) to identify and quantify proteins, their post-translational modifications (PTMs), conformations and interactions [29•]. Proteomics methods can be subdivided into two conceptual streams, top-down and bottom-up proteomics. In the top-down approach, molecular ions of the intact proteoform are generated that are then fragmented within the mass spectrometer for more extensive analyses. Bottom-up proteomics, currently the dominant paradigm for most proteomics applications, relies on the protein inference logic where peptides originating from proteins following digestion by a protease are identified and then mapped back to proteins for both identification and quantification (Figure 1a) [30]. After initially determining the mass to charge ratio of molecular ions of peptides (MS1), they are fragmented in a collision cell of a mass spectrometer and the masses and intensities of fragment ions (MS2) are determined. This information can then be used to identify the amino acid sequence of the peptide. In the last 20 years bottom-up proteomics has been dominated by two approaches, referred to as discovery and targeted proteomics [31]. In discovery proteomics the instrument is operated in data dependent acquisition mode (DDA) where the most intense peptide ions detected in MS1 scans are sequentially selected for fragmentation and MS2 analysis. Discovery proteomics can identify and quantify thousands of proteins over many conditions and samples. However, given that precursor selection for MS2 analysis occurs semi-stochastically in complex samples, this can result in a considerable number of missing values in the final quantitative data matrix if multiple samples are being sequentially analyzed, altogether impairing the reproducibility of results of cohort studies (Figure 1b). In contrast, targeted proteomics attempts to systematically acquire quantitative information on a pre-specified set of proteins by directing the mass spectrometer to deterministically acquire fragment ion signals only for a predetermined set of peptides that represent the proteins under investigation. This type of measurement primarily relies on an MS acquisition mode referred to as selected or multiple reaction monitoring, (S/MRM), or a related method termed parallel reaction monitoring (PRM), and allows the multiplexed analysis of dozens to maximally hundreds of proteins. Targeting

Figure 1


(a)
Sample preparation
Data acquisition
HPLC
Organic Solvent %
Time
MS/MS
TOF
Q1
Collision Cell
ESI
Data processing
Data analysis
(b)
MS1
Intensity
mass/charge
Intensity based selection
(Shotgun-MS)
MS2
Conditions
Proteins
(c)
Prior knowledge based selection
(S/MRM)
(d)
Full scan
(DIA/SWATH)
Protein
Peptide
Missing value
Abundance
S/MRM precursor isolation (e.g. 1 Da)
SWATH-MS window (e.g. 25 Da)
Current Opinion in Microbiology

methods show the highest accuracy and reproducibility among the mass spectrometric strategies but are not practically scalable to thousands of proteins. S/MRM requires mass spectrometric and chromatographic coordinates for the peptides of interest as prior knowledge to direct the MS data acquisition and to score the resulting data output (Figure 1c) [32,33]. In an effort to overcome the limitations mentioned above several labs have more recently developed methods based on an MS data acquisition mode referred to as data independent acquisition (DIA) in which many peptides are fragmented and analyzed in parallel. For example, SWATH-MS, an implementation of the DIA strategy, combines the strengths of both aforementioned strategies, and can systematically and reproducibly analyze up of thousands proteins across various conditions and samples (Figure 1d) [31,34]. To extract quantitative data from the highly complex and convoluted fragment ion spectra generated by DIA/SWATH-MS, prior knowledge about the fragmentation properties of specific peptides is typically employed [35]. During the past years, custom computational tools have been developed enabling chromatogram extraction [36,37], false discovery rate control [38], protein quantification [39], differential expression analysis [40,41] and study of PTMs for SWATH-MS [42].

Overall, the developed proteomics workflows can complement each other in the sense that each method is more suitable for a given biological objective. For instance, discovery proteomics (shotgun-MS) is a prerequisite for targeted proteomic studies where there is no reliable knowledge on the proteome of organism of interest. In contrast, targeting methods (S/MRM) and DIA/SWATH-MS are implemented to quantify proteins over a sample cohort consistently and reproducibly. They do require however, prior knowledge of the fragment ion spectra of the targeted peptides.

Current knowledge about the proteome of Mtb

A decade ago, shotgun-MS was the only viable mass spectrometry-based proteomic method and consequently most of the available proteomic data was in this form. In addition to providing important insights into the composition of the respective sample, the resulting fragment ion spectra provided the basis for specific measurement assays for targeting MS and DIA methods. This progression from proteome discovery to serial proteome quantification by targeting MS and DIA methods has been robustly implemented for Mtb. Specifically, Mtb is one of only a few species for which reference fragment ion spectra have been generated and made publicly accessible for proteins from every ORF of the genome [43,44••,45••]. In the following, we discuss the biological insights into the proteome of Mtb gained from different proteomic methods.

Discovery of Mtb proteomic features using DDA methods

Protein profiling of clinical isolate strains of Mtb has increased the depth of our knowledge about the Mtb proteome. Comparison of proteomic datasets on H37Rv and H37Ra, a virulent and avirulent strain of Mtb, identified 29 significant changes of membrane associated proteins, including the possible protein export membrane protein SecF and three ABC-transporter proteins, that were upregulated in H37Rv. This suggested that the bacterial secretion and transporter systems might be significant determinants for the virulence of the bacilli [46]. A similar approach applied to the membrane associated proteins of H37Rv and BCG, the vaccine strain of Mtb, revealed the significance of membrane proteins in causing the disease. Analyzing the proteome of H37Rv, H37Ra, BND and JAL strains highlighted the distinct protein expression patterns of Esx and mce1 operon proteins in the JAL and BND strain, respectively, suggesting EsxA as a potential virulence factor. Proteins MmpL4, Rv1269c, Rv3137, and SseA have been reported as major differences between the ancient and modern Beijing strains which might clarify the increased virulence and success of the modern Beijing strains [47]. SseA, a predicted thiol-oxidoreductase, together with SodA and DoxX constitute a membrane-associated oxidoreductase complex (MRC) and lack of any MRC subunit results in the defective recycling of mycothiol as a functional analog of glutathione [48,49]. The low level of SseA in the modern Beijing strains most likely results in increased DNA oxidation damage which explains the higher rate of mutation and accelerated acquisition of drug resistance compared to more ancient strains [50]. Up-regulation of enzymes responsible for long-chain fatty acid biosynthesis and HsaA implicated in steroid degradation and down-regulation of long-chain fatty acid degrading enzymes have been observed in Beijing B0/W148 strains in compared to the reference strain termed H37Rv [49]. The differential expression of 23 proteins

(Figure 1 Legend) Schematic overview on bottom-up proteomics and its three major modes. **(a)** The workflow shows the four major steps of a bottom-up proteomic study consisting of sample preparation, data acquisition, data processing and data analysis. The significant difference between various proteomic methods is the data acquisition step and associated data analysis illustrated in panel b–d. **(b)** In Shotgun-MS precursors (ionized peptides) are isolated for MS2 scan based on their intensity. The scheme depicts the selected precursors using asterisk signs. The final data matrix has a considerable amount of missing values. **(c)** S/MRM are knowledge based data acquisition methods where one has to obtain chromatographic and mass spectrometric coordinates of peptides of interest in order to define the isolation scheme. As these methods do not rely on MS1 scans for isolating precursors, they are considered as data independent acquisition methods. **(d)** SWATH-MS divides the whole range into several dozen windows and each time a subpopulation of precursors attributed to a certain mass/charge range (window) are isolated and eventually their MS2 scan is measured. In the data processing step, quantitative proteomic data is extracted based on prior knowledge.

implicated in virulence were confirmed by SRM in seven clinically relevant strains showing various degrees of pathogenicity [51]. Proteomic and transcriptional analyses also generated some insights on metabolic remodeling between different BCG strains which might be manifested by various degrees of immunogenicity and potentially vaccine efficacy [52]. The mycobacterial protein analysis of mono-infected and HIV co-infected macrophages revealed 92 significant changes which belong to various functional categories such as toxin–antitoxin (TA) modules, cation transporters and type VII (Esx) secretion systems [53].

Proteomic studies have also increased the depth of our knowledge about the significant regulatory pathways of Mtb. PhoP as a virulence factor regulates a small non-coding RNA (ncRNA) namely Mcr7 which affects the activity of the Twin Arginine Translocation (Tat) protein secretion system through TatC modulation. Consequently, the secretion of BlaC and the antigen 85 complex (Ag85), a key player in the pathogenicity, changes significantly [54]. To decipher the role of SecA2 dependent export pathway, the cell wall and cytosolic proteome of a SecA2 mutant were compared to the wild type introducing the association of the pathway with DosR regulon and the Mce1 and Mce4 lipid transporters [55]. Proteomic analyses of culture filtrate on Mtb revealed EsxG and EsxH, secreted co-dependently, facilitate the secretion of several members of the proline–glutamic acid (PE) and proline–proline–glutamic acid (PPE) protein families such as PE5 [56•].

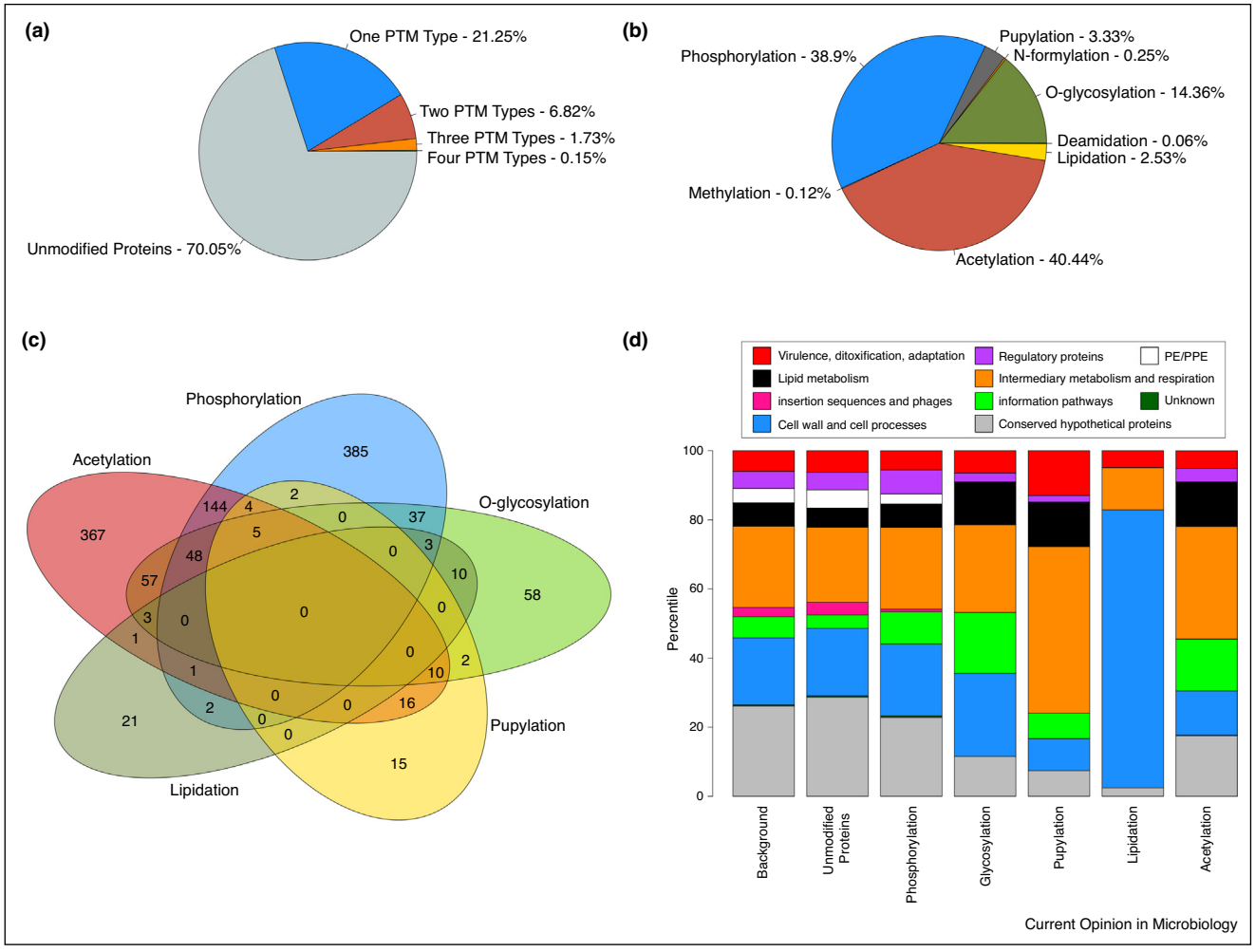
Drug resistant strains of Mtb are a growing problem for healthcare systems and have been investigated using proteomic methods. Bedaquiline (BDQ), approved for the treatment of multidrug-resistant TB (MDR-TB), inhibits ATP synthesis inducing a bacteriostatic state for 3-4 days after drug exposure. The induction of the DosR regulon as well as the activation of ATP-generating pathways promote bacterial viability during this initial drug exposure, explaining in part why BDQ is more effective when the bacilli have access to only non-fermentable energy sources such as lipids [57]. Studying the proteome of ofloxacin (OFX) resistant strains showed fourteen proteins up-regulated in respect to the OFX susceptible strains. Further docking analysis on four of the proteins elucidated conserved motifs and domains interacting with OFX as a second-line drug against MDR-TB [58]. A study showed that the abundance of several proteins responsible for the maintenance of cell-envelope permeability barrier changed significantly in Mtb exposed to thioridazine. Thioridazine increases cell-envelope permeability and thereby facilitates components uptake [59].

The term post translational modification (PTM) refers to the covalent modification of proteins by functional groups such as phosphorylation which result in different proteoforms which may represent different functional states of the protein. Although the mechanistic importance of PTMs in mammalian systems has been extensively reported, the roles of PTMs in pathogenic bacteria, such as Mtb, are not well explored. To date, eight types of protein modifications have been detected in the Mtb proteome. Our meta-analyses show that at least 30% of the Mtb proteins undergo one or more type of PTM (Figure 2a). Acetylation [60] and phosphorylation [61–65] are the most common modifications in Mtb. Moreover, O-glycosylation [66,67], pupylation (prokaryotic ubiquitin-like protein modification) [68], lipidation [66], methylation, deamidation and N-formylation have been reported (Figure 2b and c) [69]. Our analysis also showed that specific types of modifications can be enriched in specific functional categories. For instance, the lipidated proteins are mostly implicated in the cell wall processes (Figure 2d). Serine, threonine and tyrosine phosphorylation constitute ~37%, 52% and 11% of so far identified phosphorylated proteins, respectively, which is in a similar range in comparison with other bacteria (Table S1 contains a literature curated list of post translationally modified proteins detected so far in Mtb — corresponds to Figure 2). Although a tyrosine kinase with *in vivo* activity has not been identified in Mtb, some of the serine — threonine protein kinases (STPKs) revealed also tyrosine phosphorylation activities [65].

Robust and reproducible Mtb proteome profiling using DIA methods

With the recent developments in the DIA/SWATH-MS field, we can quantify ~2700 proteins of Mtb at semi-high-throughput (14 samples/day) consistently from 1 μg of total peptide mass over many samples and conditions. To support the quantitative measurement of the Mtb proteome by DIA/SWATH-MS, an Mtb proteome spectral library has been generated, validated and made publicly accessible [45••]. It contains 97% of the annotated Mtb proteins and has paved the way to study the Mtb proteome under many different conditions. According to shotgun-MS and deep RNASeq experiment, we presume that 3488 proteins are expressed in Mtb cumulatively and that protein concentrations range from 0.1 to 1000 fmol/μg (10–44 632 estimated protein copies per cell), spanning four order of magnitude. GroEL1/2, MihF, GroES and Tuf are the most abundant proteins. Furthermore, 29 previously unannotated proteins have been identified by MS-based proteomics which emphasizes that the genome annotation of Mtb still needs to be further refined [45••]. In a prototypical study, the absolute protein concentrations of the Mtb proteome and its reorganization after exposure to hypoxia was determined in a time course experiment. The results showed that whereas ribosomal proteins remain largely unchanged, products of DosR regulon genes were strongly induced to constitute 20% of the cellular protein content during dormancy. A quarter of 631 differentially expressed

Figure 2



Meta-analysis on modified proteins of *Mycobacterium tuberculosis*. **(a)** Overview on modified proteins. **(b)** Distribution of modified proteins according to their modification types. **(c)** Venn diagram shows overlap of the modified proteins with each other. **(d)** Contribution of modified proteins in various functional categories.

proteins had metabolic functions and 80% of them constituted connected metabolic pathways with at least four enzymes [44••].

Prospective directions

Systems biology approaches have de-convoluted many molecular mechanisms in various organisms. Given the functional significance of proteins in biological systems, MS-based proteomics has become a dominant method for systematic studies at high throughput. In the case of Mtb, a recently developed complete proteome spectral library has facilitated proteomic measurements of the bacilli using DIA/SWATH-MS and any targeted method. SWATH-MS is able to quantify the bulk of the Mtb proteome at semi-high-throughput reproducibly and systematically across samples and conditions and such a deep and robust protein profiling brings new biological questions to the realm of feasibility. For instance, more than two thirds of the transcription factors in Mtb can be quantified directly from the whole cell lysate which would warrant various transcriptional analyses in order to understand how transcription factors orchestrate the Mtb transcriptional regulatory network in different conditions. Since Gagneux and colleagues recently showed that even Mtb strains belonging to the sublineages of lineage 4 reveal different phenotypic traits, H37Rv as a member of that lineage cannot be the representative of Mtb clinical isolates anymore. Extending the Mtb research from H37Rv to the clinical isolates opens new opportunities for various types of studies. For instance, the GWAS approach can be portable to other biomolecular layers of a cell such as the proteome to characterize the consequence of natural genetic diversity on the proteome and eventually various phenotypes. These types of

studies, already implemented for other organisms, might manifest molecular mechanisms underlying infection, survival and persistence of Mtb.

Moving from *in vitro* research to the more physiologically relevant environment of the host should be an essential long-term goal in the field of infectious diseases. Infected cell lines or primary cells, and mouse models can be considered as intermediates toward the analysis of clinical human tissue samples. The proteomic technologies described here also support in principle the analysis of more complex samples that would include both bacterial and host components, however, quantifying pathogens' proteins following infection will be challenging due to the dilution effect of bacterial proteins into the significantly more complex proteome of the host. The significance of such studies underlines the need for developing new appropriate workflows to be able to address different biological questions through profiling clinical tissue samples more efficiently. In the ideal case, such methods would be able to quantify both pathogen and host proteomes in order to address questions in host–pathogen interactions.

We conclude that while, until recently, systems level research in the Mtb field has been dominated by genomic and functional genomic approaches, with the advent of improved methods in quantitative proteomics and increasing experience in their application to problems in Mtb biology, the field is beginning to benefit from more extensive protein measurements. We expect that protein-level information will increase the functional relevance of systems biology studies of Mtb.

Conflict of interest

None.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.mib.2017.09.013.

Acknowledgement

We would like to thank Ludovic Gillet, Tatjana Sajic and Nino Nikolovski for their intellectual contributions. This work was supported by the SystemsX.ch project TbX and the National Institutes of Health project Omics4TB Disease Progression (U19 AI106761). CN and BCC were supported by a Swiss National Science Foundation Ambizione grant (PZ00P3_161435).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. World Health Organization: *Global Strategy for Health for all by the Year 2000*. 1981.

2. World Health Organization: *Mortality and Global Health Estimates*. 2017.

3. Simeone R, Sayes F, Song O, Gröschel MI, Brodin P, Brosch R, Majlessi L: **Cytosolic access of *Mycobacterium tuberculosis*: critical impact of phagosomal acidification control and demonstration of occurrence in vivo**. *PLOS Pathog* 2015, **11**: e1004650.

4. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE *et al.*: **Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence**. *Nature* 1998, **393**:537-544.

5. Cortes T, Schubert OT, Rose G, Arnvig KB, Comas I, Aebersold R, Young DB: **Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis***. *Cell Rep* 2013, **5**:1121-1131.

6. Rose G, Cortes T, Comas I, Coscolla M, Gagneux S, Young DB: **Mapping of genotype-phenotype diversity among clinical isolates of *mycobacterium tuberculosis* by sequence-based transcriptional profiling**. *Genome Biol Evol* 2013, **5**: 1849-1862.

7. Portevin D, Gagneux S, Comas I, Young D: **Human macrophage responses to clinical isolates from the *Mycobacterium tuberculosis* complex discriminate between ancient and modern lineages**. *PLoS Pathog* 2011, **7**:e1001307.

8. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G *et al.*: **Out-of-Africa migration and neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans**. *Nat Genet* 2013, **45**:1176-1182.

9. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC *et al.*: **Variable host–pathogen compatibility in *Mycobacterium tuberculosis***. *Proc Natl Acad Sci U S A* 2006, **103**:2869-2873.

10. Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD, Gagneux S: **Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved**. *Nat Genet* 2010, **42**:498-503.

11. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW *et al.*: **High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography**. *PLoS Biol* 2008, **6**:e311.

12. •• Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, Fenner L, Rutaihwa L, Borrell S, Luo T *et al.*: ***Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages**. *Nat Genet* 2016, **48**:1535-1543.
The authors analyzed a large genomic dataset of lineage 4, the most geographically widespreed lineage, of Mtb. They found that the sublineages of lineage 4 differ in the breadth of their ecological niches called generalist and specialist. Their data revealed that the generalist sublineages have most likely originated from Europe.

13. Gygli SM, Borrell S, Trauner A, Gagneux S: **Antimicrobial resistance in *Mycobacterium tuberculosis*: mechanistic and evolutionary perspectives**. *FEMS Microbiol Rev* 2017, **41**: 354-373.

14. Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A *et al.*: **Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis***. *Nat Genet* 2013, **45**:1183-1189.

15. • Desjardins CA, Cohen KA, Munsamy V, Abeel T, Maharaj K, Walker BJ, Shea TP, Almeida DV, Manson AL, Salazar A *et al.*: **Genomic and functional analyses of *Mycobacterium tuberculosis* strains implicate ald in D-cycloserine resistance**. *Nat Genet* 2016, **48**:544-551.
They combined association analysis with a method for ampilifying signal from rare mutations in order to indentify a causative mutation of drug resistance in Mtb. They found loss-of-function mutations in Rv2780 make the bacilli restistant to D-cycloserin. Their finding was validated experimentally.

16. • Manson AL, Cohen KA, Abeel T, Desjardins CA, Armstrong DT, Barry CE, Brand J, Brand J, Jureen P, Malinga L *et al.*: **Genomic**

analysis of globally diverse *Mycobacterium tuberculosis* strains provides insights into the emergence and spread of multidrug resistance. *Nat Genet* 2017, **49**:395-402.
The largest genomic dataset of Mtb has been studied in this paper. They showed although the devirsity of Mtb is quite considerable, the pattern of the emergence of drug resistance is conserved among all the lineages. Morover, their data supports the fact that a KatG mutation (p.Ser315Thr) coffering isoniazid resistance is overwhelmingly occurred before mutations that make Mtb resistant to rifampicin.

17. Rustad TR, Harrell MI, Liao R, Sherman DR: **The enduring hypoxic response of *Mycobacterium tuberculosis***. *PLoS One* 2008, **3**:e1502.

18. Park H-D, Guinn KM, Harrell MI, Liao R, Voskuil MI, Tompa M, Schoolnik GK, Sherman DR: **Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis***. *Mol Microbiol* 2003, **48**:833-843.

19. Rustad TR, Minch KJ, Ma S, Winkler JK, Hobbs S, Hickey M, Brabant W, Turkarslan S, Price ND, Baliga NS *et al.*: **Mapping and manipulating the *Mycobacterium tuberculosis* transcriptome using a transcription factor overexpression-derived regulatory network**. *Genome Biol* 2014, **15**:502.

20. Minch KJ, Rustad TR, Peterson EJR, Winkler J, Reiss DJ, Ma S, Hickey M, Brabant W, Morrison B, Turkarslan S *et al.*: **The DNA-binding network of *Mycobacterium tuberculosis***. *Nat Commun* 2015, **6**:5829.

21. Galagan JE, Minch K, Peterson M, Lyubetskaya A, Azizi E, Sweet L, Gomes A, Rustad T, Dolganov G, Glotova I *et al.*: **The *Mycobacterium tuberculosis* regulatory network and hypoxia**. *Nature* 2013, **499**:178-183.

22. • Peterson EJR, Ma S, Sherman DR, Baliga NS: **Network analysis identifies Rv0324 and Rv0880 as regulators of bedaquiline tolerance in *Mycobacterium tuberculosis***. *Nat Microbiol* 2016, **1**:16078.
The mapped transcriptional network of Mtb was leveraged to identify which transcription factors are pivotal in orchestrating the regulatory network of Mtb in response to bedaquiline. Finally the role of Rv0324 and Rv0880 in pushing Mtb exposed to bedaquiline into a tolerant state was validated experimentally.

23. •• Chionh YH, McBee M, Babu IR, Hia F, Lin W, Zhao W, Cao J, Dziergowska A, Malkiewicz A, Begley TJ *et al.*: **tRNA-mediated codon-biased translation in mycobacterial hypoxic persistence**. *Nat Commun* 2016, **7**:13302.
This paper shows how tRNA modifications can be a significant layer of regulation in BCG exposed to hypoxia. This type of regulation more than before remind us the significance of proteomic data in any systems-level study.

24. Cortes T, Schubert OT, Banaei-Esfahani A, Collins BC, Aebersold R, Young DB: **Delayed effects of transcriptional responses in *Mycobacterium tuberculosis* exposed to nitric oxide suggest other mechanisms involved in survival**. *Sci Rep* 2017, **7**:8208.

25. Picotti P, Clément-Ziza M, Lam H, Campbell DS, Schmidt A, Deutsch EW, Röst H, Sun Z, Rinner O, Reiter L *et al.*: **A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis**. *Nature* 2013, **494**:266-270.

26. Liu Y, Buil A, Collins BC, Gillet LCJ, Blum LC, Cheng L-Y, Vitek O, Mouritsen J, Lachance G, Spector TD *et al.*: **Quantitative variability of 342 plasma proteins in a human twin population**. *Mol Syst Biol* 2015, **11**:786.

27. •• Williams EG, Wu Y, Jha P, Dubuis S, Blattmann P, Argmann CA, Houten SM, Amariuta T, Wolski W, Zamboni N *et al.*: **Systems proteomics of liver mitochondria function**. *Science* 2016, **352**: aad0189.
In this paper, the authors showed how association studies applied to a large multiomics dataset of the BxD mouse can provide biological insights. Their analysis linked several genetic variates to intermediate biomolecules such as transcripts, metabolites and proteins and eventually to phenotypes.

28. Wu Y, Williams EG, Dubuis S, Mottis A, Jovaisaite V, Houten SM, Argmann CA, Faridi P, Wolski W, Kutalik Z *et al.*: **Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population**. *Cell* 2014, **158**:1415-1430.

29. • Aebersold R, Mann M: **Mass-spectrometric exploration of proteome structure and function**. *Nature* 2016, **537**:347-355.
Two pioneers in the field of proteomics reviewed the main proteomic methods, their application and eventually prospective significance in discovering new biological insights.

30. Gillet LC, Leitner A, Aebersold R: **Mass spectrometry applied to bottom-up proteomics: entering the high-throughput era for hypothesis testing**. *Annu Rev Anal Chem* 2016, **9**:449-472.

31. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R: **Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis**. *Mol Cell Proteomics* 2012, **11** O111.016717.

32. Picotti P, Aebersold R: **Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions**. *Nat Methods* 2012, **9**:555-566.

33. Lange V, Picotti P, Domon B, Aebersold R: **Selected reaction monitoring for quantitative proteomics: a tutorial**. *Mol Syst Biol* 2008, **4**:222.

34. Collins BC, Hunter CL, Liu Y, Schilling B, Rosenberger G, Bader SL, Chan DW, Gibson BW, Gingras A-C, Held JM *et al.*: **Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry**. *Nat Commun* 2017, **8**:291.

35. Schubert OT, Gillet LC, Collins BC, Navarro P, Rosenberger G, Wolski WE, Lam H, Amodei D, Mallick P, MacLean B *et al.*: **Building high-quality assay libraries for targeted analysis of SWATH MS data**. *Nat Protoc* 2015, **10**:426-441.

36. Röst HL, Rosenberger G, Navarro P, Gillet L, Miladinović SM, Schubert OT, Wolski W, Collins BC, Malmström J, Malmström L *et al.*: **OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data**. *Nat Biotechnol* 2014, **32**:219-223.

37. Navarro P, Kuharev J, Gillet LC, Bernhardt OM, MacLean B, Röst HL, Tate SA, Tsou C-C, Reiter L, Distler U *et al.*: **A multicenter study benchmarks software tools for label-free proteome quantification**. *Nat Biotechnol* 2016, **34**:1130-1136.

38. Rosenberger G, Bludau I, Schmitt U, Heusel M, Hunter CL, Liu Y, MacCoss MJ, MacLean BX, Nesvizhskii AI, Pedrioli PGA *et al.*: **Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses**. *Nat Methods* 2017, **14**:921-927.

39. Rosenberger G, Ludwig C, Röst HL, Aebersold R, Malmström L: **aLFQ: an R-package for estimating absolute protein quantities from label-free LC–MS/MS proteomics data**. *Bioinformatics* 2014, **30**:2511-2513.

40. Choi M, Chang C-Y, Clough T, Broudy D, Killeen T, MacLean B, Vitek O: **MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments**. *Bioinformatics* 2014, **30**:2524-2526.

41. Teo G, Kim S, Tsou C-C, Collins B, Gingras A-C, Nesvizhskii AI, Choi H: **mapDIA: preprocessing and statistical analysis of quantitative proteomics data from data independent acquisition mass spectrometry**. *J Proteomics* 2015, **129**:108-120.

42. Rosenberger G, Liu Y, Röst HL, Ludwig C, Buil A, Bensimon A, Soste M, Spector TD, Dermitzakis ET, Collins BC *et al.*: **Inference and quantification of peptidoforms in large sample cohorts by SWATH-MS**. *Nat Biotechnol* 2017 http://dx.doi.org/10.1038/nbt.3908.

43. de Keijzer J, Mulder A, de Ru AH, van Soolingen D, van Veelen PA: **Parallel reaction monitoring of clinical *Mycobacterium tuberculosis* lineages reveals pre-existent markers of rifampicin tolerance in the emerging Beijing lineage**. *J Proteomics* 2017, **150**:9-17.

44. •• Schubert OT, Ludwig C, Kogadeeva M, Zimmermann M, Rosenberger G, Gengenbacher M, Gillet LC, Collins BC, Röst HL, Kaufmann SHE *et al.*: **Absolute proteome composition and dynamics during dormancy and resuscitation of *Mycobacterium tuberculosis***. *Cell Host Microbe* 2015, **18**:96-108.

This is the first paper on the application of SWATH-MS in the Mtb field. The authors analyzed the proteome of Mtb over the wayne model to understand how Mtb's proteome reorganizes in response to hypoxia. Due to the high quality and coverage of the proteomic data and their precise absolute quantification, different types of data analysis were in the realm of feasibility.

45. •• Schubert OT, Mouritsen J, Ludwig C, Röst HL, Rosenberger G, Arthur PK, Claassen M, Campbell DS, Sun Z, Farrah T *et al.*: **The Mtb proteome library: a resource of assays to quantify the complete proteome of *Mycobacterium tuberculosis***. *Cell Host Microbe* 2013, **13**:602-612.

Schubert *et al*. generated and validated the first complete proteomic library of Mtb and made this accessible to everybody. This library has paved the way to measure the Mtb proteome using targeting methods and DIA/SWATH.

46. Målen H, De Souza GA, Pathak S, Søfteland T, Wiker HG: **Comparison of membrane proteins of *Mycobacterium tuberculosis* H37Rv and H37Ra strains**. *BMC Microbiol* 2011, **11**:18.

47. Jhingan GD, Kumari S, Jamwal SV, Kalam H, Arora D, Jain N, Kumaar LK, Samal A, Rao KVS, Kumar D *et al.*: **Comparative proteomic analyses of avirulent, virulent, and clinical strains of *Mycobacterium tuberculosis* identify strain-specific patterns**. *J Biol Chem* 2016, **291**:14257-14273.

48. Nambi S, Long JE, Mishra BB, Baker R, Murphy KC, Olive AJ, Nguyen HP, Shaffer SA, Sassetti CM: **The oxidative stress network of *Mycobacterium tuberculosis* reveals coordination between radical detoxification systems**. *Cell Host Microbe* 2015, **17**:829-837.

49. Bespyatykh J, Shitikov E, Butenko I, Altukhov I, Alexeev D, Mokrousov I, Dogonadze M, Zhuravlev V, Yablonsky P, Ilina E *et al.*: **Proteome analysis of the *Mycobacterium tuberculosis* Beijing B0/W148 cluster**. *Sci Rep* 2016, **6**:28985.

50. de Keijzer J, de Haas PE, de Ru AH, van Veelen PA, van Soolingen D: **Disclosure of selective advantages in the "modern" sublineage of the *Mycobacterium tuberculosis* Beijing genotype family by quantitative proteomics**. *Mol Cell Proteomics* 2014, **13**:2632-2645.

51. Peters JS, Calder B, Gonnelli G, Degroeve S, Rajaonarifara E, Mulder N, Soares NC, Martens L, Blackburn JM: **Identification of quantitative proteomic differences between *Mycobacterium tuberculosis* lineages with altered virulence**. *Front Microbiol* 2016, **7**:813.

52. Abdallah AM, Hill-Cawthorne GA, Otto TD, Coll F, Guerra-Assunção JA, Gao G, Naeem R, Ansari H, Malas TB, Adroub SA *et al.*: **Genomic expression catalogue of a global collection of BCG vaccine strains show evidence for highly diverged metabolic and cell-wall adaptations**. *Sci Rep* 2015, **5**:15443.

53. Ganji R, Dhali S, Rizvi A, Rapole S, Banerjee S: **Understanding HIV-Mycobacteria synergism through comparative proteomics of intra-phagosomal mycobacteria during mono- and HIV co-infection**. *Sci Rep* 2016, **6**:22060.

54. Solans L, Gonzalo-Asensio J, Sala C, Benjak A, Uplekar S, Rougemont J, Guilhot C, Malaga W, Martín C, Cole ST: **The PhoP-dependent ncRNA Mcr7 modulates the TAT secretion system in *Mycobacterium tuberculosis***. *PLoS Pathog* 2014, **10**: e1004183.

55. Feltcher ME, Gunawardena HP, Zulauf KE, Malik S, Griffin JE, Sassetti CM, Chen X, Braunstein M: **Label-free quantitative proteomics reveals a role for the *Mycobacterium tuberculosis* SecA2 pathway in exporting solute binding proteins and Mce transporters to the cell wall**. *Mol Cell Proteomics* 2015, **14**:1501-1516.

56. • Tufariello JM, Chapman JR, Kerantzas CA, Wong K-W, Vilchèze C, Jones CM, Cole LE, Tinaztepe E, Thompson V, Fenyö D *et al.*: **Separable roles for *Mycobacterium tuberculosis* ESX-3 effectors in iron acquisition and virulence**. *Proc Natl Acad Sci U S A* 2016, **113**:E348-E357.

The secretory machinery of Mtb is implicated in various biological processes such as pathogenicity. This study combined genetic, proteomic and phenotypic studies to identify a few substrates of the secretion systems and their potential function.

57. Koul A, Vranckx L, Dhar N, Göhlmann HWH, Özdemir E, Neefs J-M, Schulz M, Lu P, Mørtz E, McKinney JD *et al.*: **Delayed bactericidal response of *Mycobacterium tuberculosis* to bedaquiline involves remodelling of bacterial metabolism**. *Nat Commun* 2014, **5**:6013-6018.

58. Lata M, Sharma D, Deo N, Tiwari PK, Bisht D, Venkatesan K: **Proteomic analysis of ofloxacin-mono resistant *Mycobacterium tuberculosis* isolates**. *J Proteomics* 2015, **127**:114-121.

59. de Keijzer J, Mulder A, de Haas PEW, de Ru AH, Heerkens EM, Amaral L, van Soolingen D, van Veelen PA: **Thioridazine alters the cell-envelope permeability of *Mycobacterium tuberculosis***. *J Proteome Res* 2016, **15**:1776-1786.

60. Xie L, Wang X, Zeng J, Zhou M, Duan X, Li Q, Zhang Z, Luo H, Pang L, Li W *et al.*: **Proteome-wide lysine acetylation profiling of the human pathogen *Mycobacterium tuberculosis***. *Int J Biochem Cell Biol* 2015, **59**:193-202.

61. Verma R, Pinto SM, Patil AH, Advani J, Subba P, Kumar M, Sharma J, Dey G, Ravikumar R, Buggi S *et al.*: **Quantitative proteomic and phosphoproteomic analysis of H37Ra and H37Rv strains of *Mycobacterium tuberculosis***. *J Proteome Res* 2017, **16**:1632-1645.

62. Prisic S, Dankwa S, Schwartz D, Chou MF, Locasale JW, Kang C-M, Bemis G, Church GM, Steen H, Husson RN: **Extensive phosphorylation with overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein kinases**. *Proc Natl Acad Sci U S A* 2010, **107**:7521-7526.

63. Fortuin S, Tomazella GG, Nagaraj N, Sampson SL, Gey van Pittius NC, Soares NC, Wiker HG, de Souza GA, Warren RM: **Phosphoproteomics analysis of a clinical *Mycobacterium tuberculosis* Beijing isolate: expanding the mycobacterial phosphoproteome catalog**. *Front Microbiol* 2015, **6**:6.

64. de Keijzer J, Mulder A, de Beer J, de Ru AH, van Veelen PA, van Soolingen D: **Mechanisms of phenotypic rifampicin tolerance in *Mycobacterium tuberculosis* Beijing genotype strain B0/W148 revealed by proteomics**. *J Proteome Res* 2016, **15**:1194-1204.

65. Kusebauch U, Ortega C, Ollodart A, Rogers RS, Sherman DR, Moritz RL, Grundner C: ***Mycobacterium tuberculosis* supports protein tyrosine phosphorylation**. *Proc Natl Acad Sci U S A* 2014, **111**:9265-9270.

66. Bell C, Smith GT, Sweredoski MJ, Hess S: **Characterization of the *Mycobacterium tuberculosis* proteome by liquid chromatography mass spectrometry-based proteomics techniques: a comprehensive resource for tuberculosis research**. *J Proteome Res* 2012, **11**:119-130.

67. Smith GT, Sweredoski MJ, Hess S: **O-linked glycosylation sites profiling in *Mycobacterium tuberculosis* culture filtrate proteins**. *J Proteomics* 2014, **97**:296-306.

68. Tung C-W: **PupDB: a database of pupylated proteins**. *BMC Bioinform* 2012, **13**:40.

69. van Els CACM, Corbière V, Smits K, van Gaans-van den Brink JAM, Poelen MCM, Mascart F, Meiring HD, Locht C: **Toward understanding the essence of post-translational modifications for the *Mycobacterium tuberculosis* immunoproteome**. *Front Immunol* 2014, **5**:361.