MARTIN GENGENBACHER,1 JEPPE MOURITSEN,2 OLGA T. SCHUBERT,2 RUEDI AEBERSOLD,2,3 and STEFAN H. E. KAUFMANN1

1Max Planck Institute for Infection Biology, Department of Immunology, Charitéplatz 1, 10117 Berlin, Germany; 2Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Wolfgang-Pauli Strasse 16, 8093 Zurich, Switzerland; 3Faculty of Science, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

ABSTRACT The emerging field of proteomics has contributed greatly to improving our understanding of the human pathogen Mycobacterium tuberculosis over the last two decades. In this chapter we provide a comprehensive overview of mycobacterial proteome research and highlight key findings. First, studies employing a combination of two-dimensional gel electrophoresis and mass spectrometry (MS) provided insights into the proteomic composition, initially of the whole bacillus and subsequently of subfractions, such as the cell wall, cytosol, and secreted proteins. Comparison of results obtained under various culture conditions, i.e., acidic pH, nutrient starvation, and low oxygen tension, aiming to mimic facets of the intracellular lifestyle of M. tuberculosis, provided initial clues to proteins relevant for intracellular survival and manipulation of the host cell. Further attempts were aimed at identifying the biological functions of the hypothetical M. tuberculosis proteins, which still make up a quarter of the gene products of M. tuberculosis, and at characterizing posttranslational modifications. Recent technological advances in MS have given rise to new methods such as selected reaction monitoring (SRM) and data-independent acquisition (DIA). These targeted, cutting-edge techniques combined with a public database of specific MS assays covering the entire proteome of M. tuberculosis allow the simple and reliable detection of any mycobacterial protein. Most recent studies attempt not only to identify but also to quantify absolute amounts of single proteins in the complex background of host cells without prior sample fractionation or enrichment. Finally, we will discuss the potential of proteomics to advance vaccinology, drug discovery, and biomarker identification to improve intervention and prevention measures for tuberculosis.

More than 60 years ago, Swedish biochemist Pehr Edman introduced the first technique for single peptide sequencing (1). The underlying principle involves a phenylisothiocyanate reaction with the free N-terminal amino group of a given peptide. The modified amino acid is then cleaved off and identified by chromatography or electrophoresis. Further cycles of the same process allow consecutive determination of up to 30 amino acids and thus the N-terminal amino acid sequence of a polypeptide. Major drawbacks of the so-called Edman degradation are that (i) N-terminal residues of a polypeptide must be freely accessible and unmodified and (ii) disulfide bonds cannot be directly identified. Nevertheless, Edman paved the way to modern protein identification. Proteomics has come a long way and is currently in transition from pure basic research to medical application. The reasons are obvious. The genome can be viewed as the blueprint of a cell; the transcriptome encompasses the first step, transcribing parts of the genome, which is active at a given time point. The proteome, however, describes the sum of the working
parts of a cell. Thus, proteomics is the most direct platform for measuring cellular activity. Importantly, both during transcription from DNA to RNA and during translation from RNA to protein, changes occur, which can multiply the different variants of the encoding gene. These include transcription errors, epigenetic changes, and other events, as well as translation errors, posttranslational modifications such as phosphorylation, and differential modes of protein folding. These changes increase complexity markedly, thereby allowing the most direct and most precise insight into a cell.

When the whole genome sequence of the human pathogen *Mycobacterium tuberculosis* became available in 1998, ~4,000 open reading frames were identified, and for 61% of them, explicit or putative functions could be assigned based on sequence homologies (2). The *M. tuberculosis* genome/proteome has been subdivided into 10 functional categories as depicted in Fig. 1A (TubercuList v2.6; http://tuberculist.epfl.ch/) (3). A recent study that inferred protein functions based on orthology and integrated genomic context analysis and literature mining reduced the number of hypothetical/unknown open reading frames to less than 12% (4).

Soon after initial genome annotation, scientists used random whole-genome high-density mutagenesis to identify mycobacterial genes essential for survival *in vivo* (5, 6). However, so far only a minority of the characterized encoded proteins are known to be important for infection and virulence (7–12).

Proteomics aims to provide the most detailed insights into cellular processes by analyzing mature proteins, including modifications such as posttranslational processing or cleavage, which cannot be captured by genomics or transcriptomics. Furthermore, studies in numerous species and cell types have indicated that the cellular concentration of mRNA and protein encoded by the same locus do not strictly correlate and that this correlation is state specific (13, 14). First attempts toward the proteome of *M. tuberculosis* were made by analyzing bacterial fractions and culture supernatants by two-dimensional gel electrophoresis (2D-GE). While early approaches could resolve 50 to 170 proteins (15–22), improved methods including immobilized pH gradients in the first dimension of 2D-GE resulted in ~700 distinct protein spots in a single gel (19, 20), from which ~10% could be annotated (23) (Fig. 2).

Although 2D-GE maps proved to be useful for initial proteome annotation, it soon became clear that electrophoresis-based separation methods could not be further advanced due to physical resolution limits. By using liquid chromatography–tandem mass spectrometry (LC-MS/MS) for proteome analysis (Fig. 3), a much higher identification rate could be achieved. Two main LC-MS/MS strategies are applied in proteomics today: untargeted shotgun proteomics and the more recently established targeted proteomics. Compared to 2D-GE, shotgun approaches enable superior characterization of proteins located in the mycobacterial cell wall.
or membrane, which are extremely difficult to resolve by 2D-GE due to their chemical properties. Shotgun proteomics with hybrid high-resolution MS of extensively fractionated mycobacterial lysates resulted in detection of over 3,000 proteins (24–26). Tremendous technological advances in the field of MS, notably the introduction of selected reaction monitoring (SRM), opened the field of targeted proteomics. Today, such techniques allow accurate detection and quantification of up to 80% of the \textit{M. tuberculosis} proteome in un-fractionated whole-cell lysates of liquid cultures without prior fractionation or separation (25) (Fig. 2 and Fig. 3B). Given that about 80% of the genome is expressed under such conditions, almost all proteins present are detectable, although their abundance range spans at least four orders of magnitude (25) (Fig. 1B). Further layers of complexity are added to proteomics by posttranslational modifications of proteins and their organization in complexes, both highly relevant for functionality, but impossible to deduce from genome or transcriptome data.

This chapter aims at providing a comprehensive overview of state-of-the-art knowledge in mycobacterial proteomics. First, studies describing the proteomic composition of subfractions of the tubercle bacillus, such as the cell wall and cytosol, as well as secreted proteins found in culture supernatants, will be discussed. Then, various culture conditions, i.e., low pH, nutrient starvation, low oxygen tension, or nitric oxide exposure that likely reflect facets of the intracellular life of \textit{M. tuberculosis}, move into the center of interest. A series of comparative studies provide first clues to the proteome relevant for phagosomal survival and host cell manipulation during infection. Further attempts have aimed at identifying the function of some of the hypothetical \textit{M. tuberculosis} proteins and characterizing posttranslational modifications. The cutting-edge technique of SRM, combined with a public database of specific MS assays covering the entire proteome of \textit{M. tuberculosis} now allows simple detection of mycobacterial proteins in different biological backgrounds. Using these techniques, scientists are about to move one step further by quantifying absolute amounts of single bacterial proteins in a complex mixture of proteins obtained from infected host cells without prior sample fractionation or enrichment. Finally, we discuss how proteomics can advance key application-related fields of mycobacterial research, such as vaccinology, drug discovery, and biomarker identification to improve tuberculosis (TB) prevention and intervention.
COMPARATIVE PROTEOMICS

M. tuberculosis Genome and Proteome Annotation

Most MS-based proteomic strategies rely on a high-quality protein sequence database to interpret the acquired mass spectra. However, comparison between two M. tuberculosis genome annotation exercises revealed that they are still incomplete and erroneous (27). High-throughput proteomics data can be harnessed to evaluate and refine genome annotation—a strategy called proteogenomics—by providing experimental evidence for missing genes, correcting translational start site annotations, and corroborating existing open reading frames (reviewed in reference 28). Typically, the first step of a proteogenomic analysis is the six-frame translation of the genome to capture all possible translated genomic regions. Extensive MS/MS data are then searched against this translated genome database to provide evidence for thus far unannotated open reading frames. To date, several such proteogenomic analyses have been carried out to improve M. tuberculosis genome annotations (24, 25, 29, 30). While conceptually simple, proteogenomic studies are technically difficult, because the large search space generated by the six-frame translation of the entire genome challenges the error models of database search engines. Genes newly annotated by proteogenomics should therefore be treated with caution until orthogonal data, e.g., phylogenetic conservation, corroborates their existence. The online sequence annotation database TubercuList is updated several times annually with annotation corrections and new experimental evidence for existing genes and thus is a valuable reference resource.

FIGURE 3 Uncovering the M. tuberculosis proteome. (A) Most of the current knowledge on M. tuberculosis proteomics has been generated by comparison of bacterial cultures, i.e., rich medium versus hypoxia or nutrient deprivation, conditions that M. tuberculosis might experience in vivo. Subfractions such as culture supernatant, cell wall debris, or the bacterial cytosol were separated by 2D-GE and subsequently analyzed by MS techniques. (B) Study of the M. tuberculosis proteome during infection remained difficult: Due to the overwhelming protein abundance of the host as compared to the pathogen, enrichment for bacterial fractions was required prior to analysis by 2D-GE and shotgun MS. With the availability of the complete proteome libraries for M. tuberculosis (25) and the human host (U. Kusebauch et al., in preparation), SRM will allow simultaneous proteome analysis of the pathogen and the host in complex mixtures. doi:10.1128/microbiolspec.MGM2-0020-2013.f3
for genome and proteome annotations (3, 31). The TB database (www.tbdb.org) is an equally helpful source for genome annotation and much other useful information as described elsewhere in this volume (10).

**The M. tuberculosis Secretome**
The driving forces in *M. tuberculosis* secretome research are the hunts for immunodominant antigens, biomarkers, and new drug targets (20, 30, 32, 33). Unlike most microbes, the intracellular pathogen *M. tuberculosis* has evolved several strategies to actively manipulate infected host cells and counteract their defense strategies (34). To discover proteins involved in these processes, scientists first analyzed mycobacterial culture supernatants. While very early 2D-GE studies identified only a few proteins by N-terminal sequencing (35) or immunodetection (19, 36), improved electrophoresis resolution has identified several hundred distinct protein spots in gels, of which roughly 38 could be identified by microsequencing and immunodetection (37). The detection rate could be improved using MS-based techniques to detect protein spots obtained from 2D-GE (32). It must be kept in mind when analyzing secretome data that dying bacteria in a culture might “contaminate” supernatant fractions with intracellular proteins while they lyse.

A series of studies compared *M. tuberculosis* and the vaccine strain bacille Calmette-Guérin (BCG), the attenuated derivative of *Mycobacterium bovis*, the causative agent of bovine TB (38–40). Two secreted major virulence factors, ESAT-6 and CFP-10, were among the differential protein spots observed. The genes of both proteins are located in the region of difference 1 (RD1), which encodes a type VII secretion system limited to pathogenic mycobacteria and therefore absent in BCG (41). Further secretome differences include the ESAT-6-like proteins Rv1198 and Rv1793 as well as the acetyl-CoA-acetyltransferase Rv0243. Altogether, a set of potential antigens for vaccine development was identified (40). In a recent study a combination of 2D-GE and shotgun proteomics identified 1,176 proteins from culture filtrates of exponential phase and nutrient-starved cultures (42).

**Proteomics of the Mycobacterial Cell Wall and Cytosol**
The hallmark of *Mycobacterium* species is a thick and waxy cell wall (43). Its unique architecture renders the cell envelope very rigid and extremely impermeable, protecting the pathogen from dehydration and rendering it resistant to conventional antibiotics. Recent drug penetration studies have suggested that the cell wall contributes to phenotypic drug resistance observed in nonreplicating *M. tuberculosis* (44). In contrast to genetic drug resistance, which is based on mutations located in the genes encoding target proteins, phenotypic antibiotic resistance is related to the physiological state of dormancy and is thus reversed when growth resumes (34). Proteins of the mycobacterial cell wall are of particular interest for vaccine development and diagnostics. Enzymes of the antigen 85 (Ag85) complex, a family of mycoly transferases vital for cell wall biogenesis, are among the most abundant immunodominant antigens of *M. tuberculosis* (32, 38) and thus were considered for improvement of the BCG vaccine by their overexpression and as central components of several subunit vaccines (45, 46). Although these transferases are actively directed to the cell wall compartment to fulfill their specific biological role, large quantities were found in culture supernatants (32, 38), which might have been partially extracted by detergents commonly used in mycobacterial growth media. A comprehensive study identified 306 proteins (106 unique) in the cell envelope fraction of *M. tuberculosis* (47). More optimized protein extraction methods even led to the detection of 528 cell wall proteins, of which 87 were predicted to carry a signal peptide for secretion (48). A series of studies set out to describe the cytosolic proteome and/or membrane-associated proteins (39, 47, 49–54). In conclusion, definition of the proteome of subcellular compartments *per se* strongly depends on experimental conditions and remains difficult due to the unique properties of the mycobacterial cell wall. Nevertheless, many contributing researchers have laid the foundations of our current understanding of the *M. tuberculosis* proteome and paved the way for study of models of infection (see the next section).

**Proteomics of Infection**
Upon infection, *M. tuberculosis* experiences harsh environmental changes requiring adaptation and realignment of metabolic systems. To study the processes involved, cultures of *M. tuberculosis* have been exposed to adverse conditions such as acidic pH, nitric oxide, carbon starvation, and hypoxia that likely mimic facets of the pathogen’s environment inside the host (Fig. 3A). An attractive, widely studied model for understanding dormant *M. tuberculosis* uses gradual oxygen depletion of a culture incubated in sealed glass vials (55). The proteome of hypoxic, nonreplicating bacilli in this model is characterized by elevated levels of the heat shock protein HspX (Rv2031c), bacterioferritin (BfrB, Rv3841), L-alanine dehydrogenase (Ald, Rv2780), the chaperone GroEL2 (Rv0440), putative fructose-biphosphate
alcohol dehydrogenase (Fba, Rv0363c), and translation elongation factor EF-Tu (Tuf, Rv0685). Another model of in vitro dormancy is based on carbon starvation of \( M. \) \( tuberculosis \) in physiological saline (59). Under these conditions, HspX and the hypothetical proteins Rv2557 and Rv2558 are strongly upregulated, while the secreted immune dominant antigens Mpt32 and Mpt64 and the membrane-associated trigger factor Tig are less abundant (60). Although Rv2557 and Rv2558 transcripts have been identified in human granulomas (61), both genes were dispensable for growth in the nutrient starvation model (62). HspX was also reported to be at high abundance in standing cultures and during long-term stationary phase, probably because both conditions reflect elements of both in vitro dormancy models (63, 64). More specifically, settled bacilli might experience low oxygen levels at the bottom of a culture flask, or organisms may face starvation inside large aggregates due to reduced oxygen/nutrient penetration. In line with these findings, HspX remained unchanged during \( ex \) \( vivo \) infection of host cells, which is a nonhypoxic system (65), while other results indicate elevated abundance of HspX under comparable conditions (66). The current data on the role of HspX warrants further investigation.

Employing isotope-coded affinity tagging techniques, the shift-down to dormancy by gradual oxygen depletion through two distinct phases of nonreplication was further dissected (67). Most differences observed were related to degradation and energy metabolism. A significant overlap of differentially expressed proteins exists in hypoxic nonreplicating \( M. \) \( tuberculosis \) and in the attenuated vaccine strain \( M. \) \( bovis \) \( BCG \) compared to metabolically active and replicating \( M. \) \( tuberculosis \) (38, 40, 66). It was concluded that HspX, BfrB, GroEL, Tuf, and Ald are among the proteins essential for pathogenesis.

To gain deeper insights into infection, the mycobacterial proteome during phagocytosis and its intracellular survival in the human monocytic cell line THP-1 were analyzed. In line with previous findings, elongation factor Tuf and the well-studied chaperones HspX, GroEL1, and GroEL2 were upregulated in \( M. \) \( bovis \) \( BCG \) upon phagocytosis (66). An approach involving metabolic labeling of \( M. \) \( tuberculosis \) prior to THP-1 infection revealed expression differences for 44 proteins (17). Another study identified enzymes of several metabolic pathways that are important during intracellular persistence of \( M. \) \( tuberculosis \) (65). Recent attempts aim at investigating the proteome of the tubercle bacillus \( in vivo \). In a guinea pig infection model, \( M. \) \( tuberculosis \) showed continuous variation in protein abundance for some proteins, while others remained at high abundance (68). Although each animal model has its advantages, only the nonhuman primate model closely reproduces the complexity of human TB (34). Moreover, TB infection is a very dynamic process where bacilli reside in host tissues across a wide spectrum of physiological stages, due to their adaptation to the various microenvironments. Thus, data interpretation of whole-granuloma/organ approaches is challenging. One way to improve the problem of an “averaged proteome” could be collection of samples from different granuloma types or even different sections of a single granuloma, provided they are of sufficient size and bacterial load. As discussed in the sections below, “The Mtb Proteome Library” and “Quantification of \( M. \) \( tuberculosis \) Proteins in Complex Host Backgrounds,” technological advances in MS have significantly improved detection rates and quantification accuracy for microbial proteins in complex host backgrounds, allowing us to refine our knowledge of the \( M. \) \( tuberculosis \) proteome \( in vivo \).

**FUNCTIONAL PROTEOMICS**

Expression proteomics compares protein concentrations over different conditions to infer protein activity. In contrast, functional proteomics focuses on the regulation of proteins by posttranslational modifications and protein turnover, as well as the organization of proteins into multiprotein complexes, signaling pathways, and protein networks. Ultimately these factors all contribute to the regulation of cell function, adaptation to environmental stimuli, virulence, and pathogenicity.

**Posttranslational Modifications**

To date, using data from many species, more than 300 types of posttranslational modifications have been described, which potentially increase proteome complexity by orders of magnitude (69). They control numerous processes in a cell and thus cannot be neglected. MS/MS detection of posttranslational modifications is usually based on the specific mass increase of the modified amino acid residue or by the presence of a characteristic fragment resulting from the gas phase fractionation in MS. However, due to typical substoichiometric modifications, their analysis is challenging and often requires specific strategies to achieve sensitive detection, such as enrichment or purification of analytes prior to measurement. In this section we give an overview of the most important posttranslational modifications in mycobacteria and how they have been tackled by proteomics.
Phosphorylation
Protein phosphorylation is a ubiquitous mechanism for signal transduction in all three kingdoms of life. It is a reversible posttranslational modification that is catalyzed by protein kinases and removed by protein phosphatases (70). Different types of amino acids have been shown to be phosphorylated. These include the hydroxyl amino acids serine, threonine, and tyrosine, as well as histidine and aspartate. For technical reasons, most protein phosphorylation studies to date have focused on hydroxyl amino acids. The typical MS-based phosphoproteomics workflow includes a phospho-enrichment step, because the proportion of phosphorylated peptides in a whole proteome digest is usually very low. During LC-MS/MS analysis, the phosphate groups mostly remain attached to the peptide and thus can be identified and even assigned to a specific amino acid residue (71).

Two-component systems in most prokaryotes are the predominant form of microbial phospho-based signal transduction pathways (72). Principally, they consist of a histidine kinase as the membrane-bound sensor and a corresponding response regulator. Upon a specific stimulus the sensor kinase autophosphorylates at a histidine residue and then transfers the phosphoryl group to an aspartate residue on the response regulator. Typically, the activated response regulator directly binds DNA and acts as the transcription factor to regulate gene expression. In M. tuberculosis, 11 of these phospho-relay systems have been identified (2). However, because of the acid-labile nature of histidine and aspartate phosphorolyations, they are currently out of reach of conventional MS-based proteomic workflows.

The M. tuberculosis genome encodes not only 11 two-component systems, but also an equal number of serine/threonine protein kinases (STPKs)—9 transmembrane and 2 soluble proteins—and 1 secreted tyrosine protein kinase; phosphorylation is reversed by one of the three identified protein phosphatases (73). Phosphorylation by STPKs has mostly been associated with eukaryotic organisms and has only recently been appreciated in prokaryotes. To date, the most comprehensive site-specific phosphoproteomic study of M. tuberculosis included more than 150 samples obtained from different growth and stress conditions (74). This analysis yielded 516 phosphorylation events in 301 phosphoproteins.

Pupylation
Pupylation is a recently discovered M. tuberculosis protein modification process (75). Specifically, covalent attachment of the small protein, prokaryotic ubiquitin-like protein (Pup), to a lysine residue of the target protein acts as a proteasomal degradation signal, similar to its eukaryotic counterpart ubiquitin (76). It remains to be determined whether pupylation has other, degradation-independent, regulatory functions. Three studies have employed an MS-based approach to search for potential target proteins of pupylation and identified partially overlapping sets of ~50 target proteins each (77–79). However, these studies have used expression of His-tagged Pup from an exogenous promoter to enrich pupylated proteins prior to MS analysis and thus do not precisely reflect true conditions. Functional puplyome studies where tagged Pup is solely expressed from its native promoter have yet to be performed.

Acetylation
N-Acetylation is a common posttranslational modification that occurs at lysine residues or at the N-terminus of proteins (80). Proteins with acetylated lysine residues may exhibit alterations in protein stability, interaction, localization, and function. Like many other posttranslational modifications, acetylation causes a shift in 2D-GE that can be used to distinguish modified proteins. Acetylation of ESAT-6 has, for instance, been shown to result in differential binding of CFP-10 (81). Acetylation can also be studied by LC-MS/MS, where acetylated peptides are enriched using specific antibodies and then identified by MS exploiting the characteristic mass shift of the modified amino acid (e.g., reference 82).

Lipidation
Covalent lipid modification of proteins allows the anchoring of proteins to the hydrophobic membrane. After attachment of a lipid residue and cleavage of the signal peptide, also called lipobox, the mature bacterial lipoprotein contains a diacylglycerol moiety at its N-terminal cysteine (83). The M. tuberculosis genome possesses 99 putative lipoproteins (84).

Glycosylation
The covalent attachment of carbohydrates to proteins is relatively rare in bacteria. However, during TB infection, human immune cells recognize not only mycobacterial lipoarabinomannans as antigens, but also glycosylated proteins, such as the secreted and cell surface protein Apa, highlighting the importance of mycobacterial glycoproteins (85–90). Proteomic analysis by 2D-GE and MS has led to the identification of 41 putative glycoproteins in M. tuberculosis by using ConA lectin affinity capture to enrich mannosylated proteins in M. tuberculosis culture filtrates (91). MS-based proteomics has also been applied
to identify specific O-mannosylation sites, for instance, on an isolated culture filtrate protein (FasC) of M. smegmatis in a study on bacterial protein-O-mannosylating enzyme (92). Furthermore, glycosylation sites for an additional 13 M. tuberculosis culture filtrate proteins were recently reported (93).

**Function and Interaction of Individual Proteins and Complexes**

Understanding protein-protein interactions of M. tuberculosis is important (i) because complex formation is required for proteins to transmit cellular signals, (ii) because most biochemical functions are catalyzed by protein complexes, and (iii) because interactions with host proteins are involved in host adaptation and virulence. Essential proteins in regulatory networks of M. tuberculosis and proteins directly interacting with the host are therefore potential drug targets (94, 95).

The knowledge base of M. tuberculosis protein-protein interactions is still incomplete due to the difficulty in engineering M. tuberculosis strains that express tagged bait proteins and limitations of existing technology for their analysis. Conventional yeast two-hybrid genetic screens have limited value for the study of mycobacterial interactions with their human host, due to the yeast-specific cellular machinery. Yet the system has been used as a rough screen upstream of more accurate bacterial pull-down experiments aiming to describe bacterial protein complexes (96–98). Co-immunoprecipitation (99), split protein sensor systems (100), and custom-designed shuttle vectors (101) are targeted and more accurate, but throughput is lower, and they cannot always deal with host-pathogen interactions, where computational prediction algorithms remain the preferred option today (102).

While studies of bacterial protein complexes and protein interactions with the host are challenging, recent findings in host-pathogen interaction experiments have provided significant new insight into the life cycle of M. tuberculosis in spite of the technological limitations. Upon first encounter with immune cells such as macrophages, M. tuberculosis secretes numerous soluble proteins including virulence factors which prime potential host targets (103). For instance, the abundantly secreted M. tuberculosis protein ESAT-6 directly binds to human TLR2. Here, it inhibits IRAK4-MyD88 and activates NFκB and Akt signaling (104). ESAT-6 is a known substrate of the ESX-1 secretion system, which is deleted from avirulent strains, such as M. bovis BCG (105). The proinflammatory action of ESAT-6 is further illustrated by its ability to cause endothelial secretion of matrix metalloprotease 9 (MMP9) in zebra fish. MMP9 has enzymatic activity required for the development of the hallmark inflammatory granulomatous tissue reaction in the host (106). Next, intracellular invasion is mediated by phagocytosis or macropinocytosis in an Mce protein family–dependent manner, involving direct molecular interaction with the host, although the targeted host proteins are still unknown (107, 108).

A selection of specific or redundant protein-protein interactions between M. tuberculosis and its host dictate the fate of the pathogen inside the phagosome (109, 110). To ensure survival in the macrophage, M. tuberculosis blocks fusion between the early phagosome and lysosome of the host macrophage. The underlying mechanism of action is complex, and at least three essential mycobacterial protein kinases—PknG, SapM, and PtpA—are involved that directly interact with host proteins (12, 111, 112). The further development of infection is unclear, but it was recently demonstrated that M. tuberculosis escapes into the cytosol in an ESAT-6-dependent manner in later stages of macrophage infection, eventually resulting in host cell death. Necrosis leads to inflammation and allows M. tuberculosis to escape and infect a new host cell (113). The host protein interaction partner of ESAT-6 at this stage remains unknown.

High-throughput technologies for the study of protein-protein interactions between the pathogen and host are emerging (114, 115). These methods require preselection of interaction partners. One of these approaches involves bait proteins attached to one end of a tri-functional small molecule cross-linker. Intact human host cells are incubated with the bait-linker, which contains a reactive group that reacts with the N-linked sugar moieties of the target proteins on the host cell surface. In this way the covalently attached linker accurately reports specific host-bait protein interactions. The tri-functional cross-linker contains a biotin tag with which the target proteins can be purified for identification by LC-MS/MS (115). To date, only two outer membrane proteins, OmpA and MspA, that could be targets for this approach have been identified (116–118). Discovery of host-pathogen protein-protein interactions required for M. tuberculosis invasion of macrophages therefore remains limited to a computational prediction approach (119).

**Protein Turnover**

Protein turnover of a cell refers to the biochemical dynamics of protein synthesis and degradation. Mycobacteria regulate the abundance of many of their proteins at the transcriptional level in response to stress encountered by the bacillus (120). However, the overall
correlation between transcript and protein abundances remains poor (13, 65, 121), and protein turnover is an important aspect to explain the discrepancy between transcriptional and protein levels.

To study protein turnover, bacterial cultures are typically transferred from standard growth media to a growth media supplemented with stable isotope-labeled carbon sources or amino acids, which are then incorporated into de novo synthesized proteins. Subsequently, LC-MS/MS can be used to determine the relative abundance of old and de novo synthesized proteins, reflecting the protein turnover rate (122). The first in vitro turnover study using conventional $^{13}$N-labeled broth for the growth of M. tuberculosis revealed that M. tuberculosis regulates its protein turnover dramatically and in a stress-specific manner (123). To regulate protein degradation, mycobacteria have two known systems: the essential Clp proteases that actively degrade mistranslated proteins of M. tuberculosis in vivo (124) and the ubiquitin-like protein Pup that is covalently linked to proteins, marking them for degradation by the mycobacterial proteasome. The turnover dynamics and functional importance of pupylation remain unknown (75, 125).

A nonreplicating cell state is induced in response to human hypoxic stress, such as controlled, low oxygen tension in rich medium culture. Low oxygen tension is sensed by the two-component system DosR/DosS triggering transcriptional synthesis of the DosR regulon genes (126, 127) and synthesis of proteins such as HspX (64). While the transcriptional effect only lasts for 24 hours despite sustained low oxygen tension (126), in a standing culture hypoxia model, protein expression of the DosR regulon genes in M. bovis BCG is sustained over days (23), suggesting reduced turnover rather than continuous synthesis. Compared to continuous high synthesis and degradation, reduced turnover is a more energetically efficient way for the bacillus to protect itself via stress-induced proteins in the nonreplicating state.

Typical protein turnover experiments using stable isotope labeling of proteins are time-consuming and expensive. Therefore, more efficient analytical proteomics tools that consume little sample material, combined with high throughput, robust reproducibility, and sufficient analytical depth are needed. In the next section, we describe some promising examples.

**TARGETED PROTEOMICS**

The examples described thus far emphasize that robust identification and quantification of proteins and their modifications are critical to our understanding of the physiology of M. tuberculosis and the pathology of TB. 2D-GE coupled with MS analysis, as well as shotgun/discovery proteomics, are currently the most widely used techniques for qualitative and quantitative measurements of the M. tuberculosis proteome. As described above, shotgun proteomics allows the detection of hundreds to several thousands of proteins in a single run. However this focus on high proteome coverage leads to some curtailments in reproducibility, quantitative accuracy, and sample throughput (128). The targeted MS technique SRM alleviates limitations by focusing the MS analysis on a defined set of proteins of interest (129, 130). In SRM mode, the instrument is instructed to monitor predefined combinations of peptide precursor and fragment ions, so-called transitions, continuously over time. The optimal transitions, together with the chromatographic retention time of the peptides, have to be determined for each protein of interest prior to experimentation. An emerging novel approach is SWATH-MS, which provides targeted data extraction of MS/MS spectra generated by data-independent acquisition. It is comparable to SRM in terms of accuracy and consistency but unlimited in the number of proteins that can be analyzed per run (131, 132). Importantly, for SRM and SWATH-MS, the MS coordinates of the target proteins must be known prior to their identification and quantification. The Mtb Proteome Library is a publicly accessible research resource that provides these MS coordinates for all proteins of the M. tuberculosis proteome and thus supports unbiased protein-based research in M. tuberculosis by targeted proteomics (25).

**The Mtb Proteome Library**

The Mtb Proteome Library provides information at three different levels of a proteomic workflow (25): First, the library provides information regarding proteome mapping by untargeted shotgun/discovery proteomics of extensively fractionated M. tuberculosis and BCG lysates (Fig. 4A). Data sets of several research groups have been compiled in a dedicated M. tuberculosis “build” in the PeptideAtlas database (www.PeptideAtlas.org), which allows querying of MS spectra and peptide and protein identification, and visualizes protein coverage. Second, the library provides SRM and SWATH-MS assays for almost all annotated M. tuberculosis proteins. Here, synthetic peptides have been used to determine the most intensely fragmented ions, as well as the chromatographic retention times of each peptide (Fig. 4B). These assays can be downloaded from the M. tuberculosis build in the
SRMAtlas database ([www.SRMAtlas.org](http://www.SRMAtlas.org)). Third, the library contains SRM data validating these assays for over 70% of the annotated proteins ([Fig. 4C](#fig4c)). All SRM signals can be browsed and inspected in the PASSEL database ([www.PeptideAtlas.org/passel](http://www.PeptideAtlas.org/passel)), permitting selection of the most specific and informative assays for proteins of interest, based on information on detectability.

**Quantification of *M. tuberculosis* Proteins in Complex Host Backgrounds**

Transcriptomics and proteomics have provided deep insights into mycobacterial responses to stress *in vitro*, but limited insights into the *in vivo* responses ([133–137](#ref133-137)). Thus, the specific adaptation mechanisms for persistence in the infected host are still unknown.
The direct quantification of mycobacterial proteins in infected human cells is complicated by the complexity of the human background proteome that spans more than seven orders of magnitude in dynamic range and the usually low multiplicity of infection. In vivo expression of mycobacterial proteins in cells and tissues has been studied primarily with antibody-based assays and green fluorescent protein fusion constructs (138, 139). While these methods show high sensitivity, they suffer from low throughput and limited ability to multiplex. Furthermore, green fluorescent protein assays cannot be applied to the study of human lesions, and few antibodies specific for *M. tuberculosis* proteins are known. To make bacteria in infected cells accessible to biochemical analysis, they were isolated from the phagosomal compartment in which they reside, and these isolates were subjected to 2D-GE-based separation for MS analysis. The success of such studies has been limited, mainly because of an excess of human proteins in the isolated phagosomes and the inherent variability of phagosomal populations (65, 66, 140).

Direct quantification of mycobacterial proteins in unfractionated human background has been challenging. *M. tuberculosis* proteins were quantified in infected macrophage cell lines (e.g., THP-1 cells) with the use of quantitative 2D-GE, but this only led to identification of the most abundant heat shock proteins in spite of high experimental multiplicity of infection (17). Classical shotgun-based protein quantification of full lysates of whole guinea pig lungs infected with *M. tuberculosis* at a high multiplicity of infection detected over 500 proteins (68). A recent study of cerebrospinal fluid from tuberculous meningitis patients using 2D-GE and spot quantification revealed differential regulation of a small set of *M. tuberculosis* proteins (141). Intriguingly, two independent studies using antibody-based techniques to detect mycobacterial proteins in TB patients indicated that secreted Ag85 protein was detectable in both serum and urine (142, 143). Furthermore, researchers studying urine from TB patients detected a number of cell membrane–associated metabolic *M. tuberculosis* proteins by shotgun MS (144). These results are encouraging because easily accessible human body fluids represent ideal specimens for the detection of *M. tuberculosis* biomarkers.

The above-mentioned studies using direct proteomic quantification of *M. tuberculosis* in human samples have provided new insight into the life and adaptation of *M. tuberculosis* in the host environment, but the robust quantification of mycobacterial proteins in the context of the host proteome remains a significant challenge. SRM is the most sensitive, selective, and reproducible LC-MS/MS-based technique currently available (128). Furthermore, targeted proteomics by SRM is easily multiplexed, and, in comparison with affinity reagent-based methods, assay development is fast and unambiguous (145). In fact, the Mtb Proteome Library resource described above contains specific SRM assays for 97% of the predicted open reading frames in the *M. tuberculosis* genome and thus supports, in principle, the quantification of essentially any *M. tuberculosis* protein in any background, including the human host. However, in reality, technical constraints, such as the dynamic range of the proteins of the human host matrix, the limited sample capacity of the chromatography columns used in LC-MS/MS measurements, and potential signal interferences and ion suppression, limit the scope of such analyses (146). It follows that any enrichment of the mycobacterial protein over the host protein background will improve bacterial protein detection. In a direct quantification workflow, the ratio of bacterial to host protein can be maximized by (i) an experimental design using a higher multiplicity of infection that does not cause unwanted host cell pathology, (ii) optimizing the cell lysis and protein extraction protocol, and (iii) specific phagosome enrichment/purification methods. An optimal sample preparation protocol, combined with the use of optimized SRM assays, has the potential to make a significant fraction of the *M. tuberculosis* proteome directly detectable in the matrix of the human host cells. If these measurements are performed with the inclusion of stable isotope-labeled internal peptide standards, the detected proteins can also be accurately quantified by the same method (129, 147).

For optimal results, the specificity of the SRM assay should be asserted in the specific host background prior to making actual measurements. There are two levels of specificity for SRM-based quantification: the peptide and the fragment ion level. Very few SRM-suited, tryptic *M. tuberculosis* peptides that are 7 to 21 amino acids in length have identical sequences in the two host species. These few peptides are predominantly derived from highly conserved central metabolic enzymes and should be excluded from SRM analyses. The fragment ion level specificity computed by the SRMCollider is depicted in Fig. 5 (148). The six most intense fragment ions of a peptide precursor of the respective protein are sufficient to select and develop a specific, scheduled SRM assay in human background for the vast majority of the proteome (25).

Direct quantification of mycobacterial proteins in infected cells and tissue inevitably captures all bacterial...
proteins in the sample, including proteins from non-replicating bacteria. The interpretation of differential expression data in infected cells is influenced by inaccuracies in bacterial quantification when comparing two or more samples. Further complications arise in tissue analysis when extracellular proteins are secreted from live or shed from dead bacteria. The gold standard bacterial quantification method is plating of colony-forming units on solid, rich medium. However, only platable, replicating bacteria are detected. Additionally, the plating method has considerable technical variation, and other methods need to be considered for enumerating bacteria in complex samples (149). In infected cells the multiplicity of infection varies with infection efficiency and replication post-infection, and mycobacterial load varies considerably in human lesions (150). It follows that protein level sample normalization is a prerequisite for robust studies in human body fluids, infected cells, and tissues. Gene expression levels are predominantly normalized to a presumably stably expressed gene like sigA or 16S RNA (133, 136, 151). However, a single reference point is outlier-sensitive, and more comprehensive strategies for protein quantification in complex host backgrounds are presumably required. These challenges notwithstanding, it can be expected that the increasingly powerful, targeted proteomics techniques will yield valuable information about the state of *M. tuberculosis* in the human host.

**IMPACT OF PROTEOMICS ON TB INTERVENTION AND DIAGNOSTICS**

Proteomics provides exciting new insights into the life cycle of *M. tuberculosis* under different conditions ranging from active metabolism and replication to dormancy. It also creates a sophisticated basis for the development of novel intervention measures against TB. This holds true for the three pillars of TB control: vaccines, drugs, and diagnostics.

**TB Vaccine Strategies**

The only TB vaccine currently in use is the live attenuated *M. bovis* strain, BCG. It protects against severe forms of childhood TB but not against pulmonary TB at all stages of life. Thus, BCG has outlived its usefulness, and new vaccines are needed for efficient TB control (152). Current vaccine design against TB follows two strategies. First, improvement of BCG by genetic modification. These vaccine candidates are currently considered for BCG replacement. Second, subunit vaccines composed of defined antigens, which are given as heterologous boosters on top of BCG prime. These vaccines comprise defined antigens. Thus far, antigen selection has been biased toward abundant antigens in broth cultures of *M. tuberculosis*. It has been proposed that the most abundant antigens of *M. tuberculosis* could benefit the pathogen rather than the host (153). As a corollary, novel strategies for vaccine-antigen discovery are warranted. Proteomics can provide clues to the discovery of better vaccine antigens. By means of proteomics, antigens expressed by *M. tuberculosis* under defined conditions with relevance for vaccine development can be identified. The following steps of experimental conditions, according to increasing complexity, could be pursued (Table 1):

- *M. tuberculosis* cultured under different conditions mimicking favorable or adverse conditions including stress, hypoxia, and nutrient deprivation.
- *M. tuberculosis*-infected host cells comprising (i) cells involved in *M. tuberculosis* control, i.e., professional phagocytes and antigen-presenting cells, notably, mononuclear phagocytes and den-
Table 1 Potential of proteomics for design of TB intervention measures

<table>
<thead>
<tr>
<th>Experimental Approach</th>
<th>Condition</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteome</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>M. tuberculosis</em> in broth culture</td>
<td>Different growth conditions: optimal, hypoxia, nutrient-deprivation, stress Drug-induced perturbations</td>
<td>Proteome of <em>M. tuberculosis</em> in cultures mimicking different infection stages Proteome of <em>M. tuberculosis</em> under influence of novel drugs</td>
</tr>
<tr>
<td><em>M. tuberculosis</em> and drug in broth culture</td>
<td><em>M. tuberculosis</em> infection of professional phagocytes, antigen-presenting cells, cellular niches</td>
<td>Proteome of <em>M. tuberculosis</em> in different cellular habitats</td>
</tr>
<tr>
<td><em>M. tuberculosis</em> in host cells</td>
<td>Phagosome, cytosol</td>
<td>Proteome of <em>M. tuberculosis</em> during subcellular compartments</td>
</tr>
<tr>
<td><em>M. tuberculosis</em> in subcellular compartments</td>
<td>Solid, necrotic, caseating lesions</td>
<td>Proteome of <em>M. tuberculosis</em> during different stages of infection (LTBI/active TB), severity of disease</td>
</tr>
<tr>
<td><em>M. tuberculosis</em> in granulomatous lesions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peptidome</td>
<td>Elution of peptides from antigen-presenting molecules MHC class II, MHC class I, unconventional antigen-presenting molecules</td>
<td>Identification of <em>M. tuberculosis</em> peptides relevant to T cell stimulation (CD4 T cells, CD8 T cells, unconventional T cells)</td>
</tr>
</tbody>
</table>

...dritic cells, as well as (ii) cells that presumably serve as protective niches for *M. tuberculosis*, such as lung Clara cells, and alveolar epithelial cells of type I and type II.

- *M. tuberculosis* residing in cellular compartments, notably in the phagosome and the cytosol, could add further information.
- *M. tuberculosis* isolated from granulomatous lesions of different types, i.e., solid granulomas, necrotic granulomas, and caseating cavitary lesions, would best reflect the *in vivo* setting of different stages of infection and disease.

In addition to proteomics, identification of peptides from *M. tuberculosis* that are generated in professional antigen-presenting cells should be revitalized. These peptides are presented to T cells, the central mediators of protective immunity. Thus, following a strategy developed decades ago (154), modern proteomics technologies could help to define the immune-relevant peptidome of *M. tuberculosis* in its completeness. Peptides are generated in different subcellular compartments of antigen-presenting cells, notably, the phagosome and cytosol. Antigenic peptides generated in the phagosome are loaded onto gene products of the major histocompatibility complex class II (MHC II) and presented to CD4 T cells. In contrast, antigenic peptides originating from the cytosol, or generated through cross-priming, are loaded on MHC class I peptides for CD8 T cell priming. These peptide epitopes are generally characterized by distinct anchor amino acids required for binding to the MHC cleft and are of ~8 or ~20 amino acids in length, for MHC class I or II, respectively (154).

After appropriate analysis and exclusion of human proteins/peptides, all *M. tuberculosis* protein antigens and peptide epitopes can be compiled and tested for their suitability as vaccine antigens.

**TB Drug Discovery**

Proteomics can guide rational drug discovery against TB (155). The increasing incidences of multidrug-resistant (MDR) TB, mounting to ~0.5 million cases annually, notification of extensively drug-resistant (XDR) TB in 84 countries globally, and the emergence of totally drug-resistant (TDR) TB urgently call for new TB drugs. Principally, the approach follows the one delineated for vaccine development. In addition, cultures of *M. tuberculosis* in broth under different conditions and in the presence of drug candidates will cause overall perturbations in the pathogen (Table 1). The available drugs against TB all target metabolically active and replicating *M. tuberculosis* organisms. Hence, dormant *M. tuberculosis* with a highly reduced metabolic and replicative activity is phenotypically resistant against these drugs. It is generally accepted that novel drugs should target not only active *M. tuberculosis*, but also dormant *M. tuberculosis*. Such drugs would not only provide alternatives for treatment of drug-resistant *M. tuberculosis*, but also reduce treatment time. Current drug discovery is strongly biased toward a linear approach of one drug targeting one unique molecule in *M. tuberculosis*. Systems biologic analysis, however, has shown that even a single perturbation starting with a single target molecule causes complex overall changes in the proteome of *M. tuberculosis*, with profound consequences for the microbe. Proteomics of *M. tuberculosis* after drug perturbation will reveal this complex network and thus provide salient information for downstream drug development. The drug discovery platform...
would also benefit enormously from the proteome maps of *M. tuberculosis* cultured under different conditions, in different host cells, and within different types of lesions as described above. Comparisons of these signatures can facilitate identification of novel drug candidates targeting *M. tuberculosis* proteins differentially expressed under defined conditions such as dormant *M. tuberculosis* as it persists during latent *M. tuberculosis* infection (LTBI) and metabolically active and replicating *M. tuberculosis* as it ravages during active TB.

Obviously, proteomics need not be restricted to the pathogen but should also include the host response. Proteome analyses of host responses to *M. tuberculosis*, targeted by drugs, will lead to a better understanding of the consequences for host cells, and the macroorganism as a whole, and could therefore predict downstream events, either beneficial or adverse to the host. Finally, proteomics of host cells, under the pressure of *M. tuberculosis* infection, will promote identification of novel drug targets in the host for treatment of TB. *M. tuberculosis* survives in host cells not only because of its robustness against aggressive defense mechanisms, but also because it actively subverts or modifies these mechanisms. Proteomics of *M. tuberculosis*-infected host cells will lead to the identification of host proteins that are differentially regulated during *M. tuberculosis* infection. By means of knockdown strategies, the relevance of these host molecules to control of intracellular *M. tuberculosis* can be revealed. It is expected that a number of signaling cascades involved (e.g., in the generation of reactive oxygen and nitrogen intermediates, in acidification of phagosomes, in phagosome-lysosome fusion, and in apoptosis/autophagy) are actively affected by *M. tuberculosis*. Such analyses can reveal targets for novel intervention strategies against TB aimed not directly at the pathogen, but indirectly affecting *M. tuberculosis* through modification of host effector mechanisms. This novel therapeutic approach combining chemotherapy and immunomodulation could open new avenues toward discovery of novel TB drug candidates against hitherto unexploited targets.

**Biomarkers for TB**

A third important area for proteomics is the design of biosignatures that allow discrimination between patients with active TB, individuals with LTBI, and uninfected healthy individuals. In the future, biosignatures that predict risk of active TB disease in LTBI subjects represent an ambitious approach, which would be of enormous value for TB control (156, 157). Proteome-based biomarker studies will likely focus on the host response rather than the pathogen. Probing the *M. tuberculosis* proteome with sera from individuals with LTBI and patients with active TB could allow identification of the antibody repertoire that is directed at all accessible *M. tuberculosis* proteins in the host and therefore lay the foundation for novel serodiagnostic tests. Although serodiagnosis of TB thus far has failed (158), novel tests comprising proteins that are reproducibly and specifically recognized by antibodies during distinct stages of infection (LTBI versus active TB disease) can be envisaged by exploiting unbiased analysis of the immunoproteome of *M. tuberculosis*. Following the experimental settings outlined above, namely, *M. tuberculosis* under different culture conditions, in different host cells, and derived from different types of lesions, proteomics of *M. tuberculosis* will create an in-depth overview of up- or downregulated proteins and thus provide guidelines for proteome-based *M. tuberculosis* diagnosis, be it in serum, sputum, or urine. High specificity of such a diagnostic assay would require selection of proteins specific for *M. tuberculosis* and absent in the vaccine BCG and in nontuberculous mycobacteria. In addition, probing of these *M. tuberculosis* proteomes with serum antibodies from LTBI and TB patients generated under different conditions will provide helpful information that can lead to further refinement of antibody-based diagnostics for TB.

**ACKNOWLEDGMENTS**

We thank Mary Louise Grossman and Diane Schad for outstanding editorial support and excellent graphic design, respectively.

This work received financial support from the European 7th Framework Program SYSTEMTB (HEALTH-2009-2.1.2-1-241587).

**REFERENCES**


Mycobacterium tuberculosis in the Proteomics Era


