DISSERTATION

EVALUATION AND CHARACTERIZATION OF TUBERCULOSIS SERODIAGNOSTIC BIOMARKERS

Submitted by

Mark Jonathan Sartain

Department of Microbiology, Immunology, and Pathology

In partial fulfillment of the requirements

For the degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2007

COLORADO STATE UNIVERSITY

November 1, 2007

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY MARK JONATHAN SARTAIN ENTITLED EVALUATION AND CHARACTERIZATION OF TUBERCULOSIS SERODIAGNOSTIC BIOMARKERS BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

COMMITTEE ON GRADUATE WORK

Norman P. Curthoys

Kenneth F. Reardon

Herbert P. Schweizer

Richard A. Slayden

Advisor: John T. Belisle

Department Head: Carol D. Blair

ABSTRACT OF DISSERTATION

EVALUATION AND CHARACTERIZATION OF TUBERCULOSIS SERODIAGNOSTIC BIOMARKERS

Detection of the antibody response has the potential to provide early and sensitive diagnosis of infection and disease, a critical element of tuberculosis control. The objective of this study was to discover and evaluate *Mycobacterium tuberculosis* antigens with protein microarray technology. Techniques were established for separation of native *M. tuberculosis* cytosol and culture filtrate proteins, resulting in 960 protein fractions that were used to generate protein microarrays. Evaluation of serological reactivity from 42 patients in three tuberculosis disease states and healthy purified-protein-derivative positive individuals demonstrated distinct reactivity profiles. Identification of antigens within the reactive fractions yielded eleven products recognized by both cavitary and noncavitary TB patients, and four proteins (HspX, MPT64, PstS1, and TrxC) specific to cavitary tuberculosis patients. Moreover, this approach identified four novel B cell antigens (BfrB, LppZ, SodC, and TrxC) of human tuberculosis, and the seroreactivity of these antigens was validated with more conventional techniques.

SodC is one of two superoxide dismutases produced by *M. tuberculosis* and was previously shown to contribute to virulence. SodC is also a putative lipoprotein

and like other sec-translocated mycobacterial proteins it was suggested to be modified with glycosyl residues. The objective of this study was to define the sites and extent of SodC glycosylation. An approach was developed that combined site-directed mutagenesis, lectin binding, and mass spectrometry. Six O-glycosylated residues within a 13 amino acid region were identified near the N-terminus. Each residue was modified with one to three hexosyl units, and the most dominant SodC glycoform was modified with nine hexosyl units. In addition to O-glycosylation of threonine residues, this study provides the first evidence of serine O-glycosylation in mycobacteria. When combined with bioinformatic analyses, the clustering of O-glycosylation appeared to occur in a region of SodC with a disordered structure and not in regions important to the enzymatic activity of SodC. The use of recombinant amino acid substitutions to alter glycosylation sites provided further evidence that glycosylation influences proteolytic processing and ultimately positioning of cell wall proteins.

Mark Jonathan Sartain
Department of Microbiology,
Immunology, and Pathology
Colorado State University
Fort Collins, CO 80523
Fall 2007

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. John Belisle, for his full support of my graduate research and education. Dr. Belisle encouraged independent, critical thinking, provided direction when needed, and guided my development as a scientist. I would also like to thank my graduate committee members, Dr. Norman Curthoys, Dr. Kenneth Reardon, Dr. Herbert Schweizer, and Dr. Richard Slayden, for their participation, enthusiasm, and valuable insight regarding my research. I thank Dr. Suman Laal and Dr. Krishna K. Singh for a very productive collaboration. I would also like to thank the Mycobacteria Research Laboratories. I have enjoyed the interactions within this research group and am grateful for the friendships that have developed.

I especially want to thank my entire family for their support. I thank my parents, Joe and Joyce, for encouraging me to pursue only that which makes me happy. I thank my brother, Matthew, for not allowing me to take myself too seriously. Finally, I thank my wife Ashlyn for her unending love, encouragement, and patience during these challenging years.

DEDICATION

To my wife Ashlyn:

for changing my life

for the better

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

2-D	two-dimensional
AIDS	acquired immunodeficiency syndrome
AFB	acid-fast bacilli
AG	arabinogalactan
Ala	alanine
AIEX	anion exchange
BCA	bicinchoninic acid
CE	capillary electrophoresis
CF	culture filtrate
CFP	culture filtrate proteins
ConA	Concanavalin A
CR	complement receptor
CV	column volume
DC-SIGN	DC-specific intercellular adhesion molecule-3-grabbing nonintegrin
DOTS	Directly observed therapy, short-course
DTH	delayed-type hypersensitivity
E:S	enzyme to substrate
ELISA	enzyme-linked immunoadsorbant assay
ESI	electrospray ionization
GAS	glycerol-alanine-salts
GS	glutamine synthetase
HBHA	heparin-binding hemagglutinin
HIV	Human Immunodeficiency Virus
HIV-TB+	HIV-noninfected, TB-infected
HIV+TB+	HIV-coinfected TB
HK	histidine kinase
HUPO	Human Proteome Organization
IFN-γ	Interferon-γ
LB	Luria-Bertani
LC	liquid chromatography
Ig	immunoglobulin
IL-12	interleukin-12
IMAC	immobilized metal affinity chromatography
LAM	lipoarabinomannan
LC	liquid chromatography
LM	lipomannan
LMW	low molecular weight

LRSITRD	Lala Ram Sarup Institute of Tuberculosis and Respiratory Diseases
Lys	lysine
MA	mycolic acids
MALDI	matrix-assisted laser desorption/ionization
mAGP	mycolic acid-arabinogalactan-peptidoglycan
MHC	major histocompatibility complex
MDR-TB	multi-drug resistant TB
MMR	macrophage mannose receptor
MRM	multiple reaction monitoring
MS	mass spectrometry
MS/MS	tandem mass spectrometry
Mtb	*Mycobacterium tuberculosis*
MTBC	Mtb complex
MudPIT	multidimensional protein identification technology
OADC	oleic acid-dextrose catalase
ORF	open reading frame
NO	nitric oxide
NOS2	nitric oxide synthase 2
PAGE	polyacrylamide electrophoresis
PBMC	peripheral blood mononuclear cell
PG	peptidoglycan
PIM	phosphatidylinositol mannoside
PM	plasma membrane
pMHC	peptide-MHC array
PSP	pulmonary surfactant protein
PPD	purified protein derivative
PPP	Plasma Proteome Project
PRR	Pattern Recognition Receptor
PSA	Prostate Specific Antigen
PTM	posttranslational modification
RD1	region of deletion 1
RNI	reactive nitrogen intermediates
ROI	reactive oxide intermediates
RP	reversed phase
RR	response regulator
SARS	severe acute respiratory syndrome
SD	standard deviation
SDS	sodium dodecyl sulfate
SELDI	surface-enhanced laser desorption/ionization time-of-flight
Ser	serine

SOD	superoxide dismutase
SPase	signal peptidase
TACO	tryptophan aspartate-containing coat protein
TB	Tuberculosis
TDM	trehalose dimycolate
Thr	threonine
TLR	toll like receptor
TNF-α	tumor necrosis factor α
Tyr	tyrosine
MVAMC	Manhattan Veteran Affairs Medical Center
WCL	whole-cell lysate
WHO	World Health Organization
XDR-TB	extreme-drug resistant TB

Chapter I

Literature Review Part I: Tuberculosis History, Pathogenesis, Host Response, Diagnosis, and Protein Posttranslational Modifications

1.1 Tuberculosis: Past and present

Tuberculosis (TB) is an ancient disease, believed to have affected human populations over a wide geographic distribution ever since the emergence of *Mycobacterium tuberculosis* (*Mtb*) as a human pathogen, a speciation event estimated to have occurred between 15,000 and 20,000 years ago (147, 272). Direct evidence for human TB infection in ancient times was provided with numerous identifications of mycobacterial DNA extracted from lung and bone lesions from exhumed, mummified human remains (238, 286, 287, 325), with the oldest sample dating back to 5,400 years ago (59). The Greek scholar Hippocrates (460 BC) provided some of the first medical descriptions of TB-like symptoms (101); however, it was not until the nineteenth century, in the midst of a major North American and European TB epidemic, that the French surgeon Jean-Antoine Villemin provided evidence that TB was an infectious disease (101). Yet widespread societal acceptance of the contagion theory did not begin until the German bacteriologist Robert Koch in 1882 identified the TB bacterium as the causative agent of the disease (101). This discovery coincided with the rapid decrease of disease incidence in the Western World from 1840 to 1960, a decline attributed to numerous factors including improved standards of living, segregation of patients, the pasteurization of milk, changes in public healthy policies, and finally the arrival of the antibiotic era (101).

The antibiotic revolution, the observed decline of TB in the industrialized nations, and advances in disease control such as the eradication of smallpox, prompted many leading scientists and physicians in the 1960s and 1970s to predict the eradication of TB by the year 2000 (101). However, this sense of optimism ended with the rapid resurgence of TB in New York City and other urban areas in the 1980s to early1990s. These isolated epidemics ultimately helped to bring attention to the global rise in TB incidence (101). This harsh realization prompted The World Health Organization (WHO) to declare TB a global health emergency in 1993, and to initiate the DOTS (directly observed therapy, short-course) program that has brought considerable success towards controlling TB in many countries (1). In fact, the most recent 2007 WHO report suggests global TB may be on the threshold of decline, with incidence rates stable in 2004 and declining in some regions in 2005 (1). This progress is attributed to the WHO Stop TB Strategy that includes the DOTS program, currently applied in 187 countries and covering 89% of the world's population (1).

Although preliminary signs of progress are encouraging, worldwide TB is at epidemic proportions. The WHO estimates that there were 8.8 million new TB cases in 2005 (1). TB is responsible for nearly two million deaths annually, accounts for one in four adult preventable deaths, and is second only to human immunodeficiency virus (HIV) as the leading cause of death by an infectious disease (http://www.wpro.who.int/media_centre/fact_sheets/fs_20060829.htm). The situation is worsening in places like sub-Saharan Africa, where the spread of HIV results in increased TB disease progression and mortality. In fact, TB is the leading opportunistic infection of acquired immunodeficiency syndrome (AIDS) patients, and causes 13% of AIDS-

related deaths globally (http://www.who.int/tb/challenges/hiv/faq/en/index.html). Further threatening progress in TB control efforts is the emergence of multi-drug resistant (MDR) -*Mtb*, accounting for 425,000 new TB cases in 2005 (http://www.who.int/mediacentre/news/releases/2006/pr24/en/), and especially frightening are recent outbreaks of extensively-drug resistant (XDR) -TB, where the global prevalence is not yet known (1, 49, 133). In light of these challenges, the development of novel drugs, diagnostics, and effective vaccines will be required for continued progress in controlling the TB crisis.

1.2 The bacillus

Mycobacterium spp. are aerobic, rod-shaped bacteria belonging to the family *Mycobacteriaceae* of the order *Actinomycetales*. *Mycobacterium* spp. have a high DNA G+C content (62-70%) and possess long chain mycolic acids in their cell wall that provide resistance to acid decolorization in staining procedures and distinguish this genus *Mycobacterium* as acid-fast. *Mycobacterium* spp. can be divided into fast-growing (one to four hour doubling time) and slow-growing (12 to 24 hour doubling time) species. The fast-growers, such as *Mycobacterium smegmatis* and *Mycobacterium phlei*, are found in the environment and are generally considered nonpathogenic. Slow-growing species are typically pathogenic and are well-adapted to mammalian hosts, and include *Mycobacterium leprae*, the causative agent of leprosy, and members of the *M. tuberculosis* complex (MTBC). The MTBC is composed of *Mtb*, *Mycobacterium africanum*, *Mycobacterium canettii*, *Mycobacterium bovis*, and *Mycobacterium microti*. Each share identical 16S rRNA sequences (28). *M. africanum, M canettii*, and primarily *Mtb* cause TB in humans, while *M. microti* causes disease in voles. *M. bovis* causes

disease in a wide variety of mammalian species including humans, and the live vaccine strain bacille Calmette-Guerin (BCG) is an attenuated derivative of *M. bovis*. Recently, comparative genomic analyses demonstrated that each member of the MTBC likely co-evolved from a common ancestor, and contrary to previous dogma *M. bovis* evolved from *Mtb* (35). The most widely-used laboratory strain of *Mtb* is H37, isolated in 1905, (6), and further dissociated into the variants H37Ra and H37Rv in 1934 (274). In 1998 the complete *Mtb* H37Rv genome sequence was determined (51), and was soon followed by the sequencing of several other *Mycobacterium* spp. genomes. This information has made an enormous impact on TB research efforts in the last decade.

It is necessary to briefly address the unique physiology of the tubercle bacillus. The cellular envelope has been a subject of intense investigation, and the molecular structures of its components are well defined. In addition to providing an especially effective physiological barrier, the diverse array of cell envelope molecules has been demonstrated to contribute to pathogenesis in a variety of manners (63, 64), some of which will be discussed later in this chapter.

The *Mtb* envelope begins with the plasma membrane (PM), a typical bacterial phospholipid bilayer uniquely possessing tuberculostearic acid and phosphatidylinositol mannosides (PIMs). Extending outward from the PM is the mycolic acid-arabinogalactan-peptidoglycan complex (mAGP), composed of peptidoglycan (PG) that is covalently linked to arabinogalactan (AG), which is itself esterified to mycolic acids (MA) (57). The PG is thought to be relatively thin, and is of the A1γ type (247), analogous to *Escherichia coli* PG. However, the PG is unusual in that it contains a mixture of N-glycolylated and N-acetylated muramic acids (175), and the peptide

carboxylic acids are often amidated (149). AG is covalently attached to PG via an α-Rha-(1→3)-α-GlcNAc-(1→P) disaccharide linker unit through a phosphodiester bond at the C6 position of muramic acid residues (182). AG is a tripartite structure consisting of the linker region, an alternating 1→5,1→6 β-D-galactofuran, and a branched D-arabinofuran (62). Attached to a portion of the terminal arabinosyl residues are the MA. MA are α-alkyl, β-hydroxy fatty acids 70 to 90 carbons in length and are the longest fatty acids found in nature (57). MA are thought to run perpendicular to the PM and form a lipid barrier likely responsible for the low permeability to many molecules, thereby contributing to drug resistance. However, the spatial arrangement of these structures is a matter of speculation, and is largely inferred from electron microscopy using freeze-substitution methods (57).

The structurally related PIMs, lipomannan (LM), and lipoarabinomannan (LAM) are acylated polysaccharides noncovalently associated with the cell wall (57). PIMs 1 to 6 are phosphatidylinositol molecules modified with one to six mannose residues, with PIM2 and PIM6 being the most abundant (202). PIMs can additionally be acylated at the 3-position of the inositol and at the 6-position of the 2-linked single mannose (107). PIM4 can be thought of as the building block upon which the mannan core is elongated, as PIM6 contains two α(1→2)-linked mannoses not found in the backbone of LM or LAM (148). LM basically consists of PIM4 modified with a α(1→6)-linked mannan containing branching α(1→2)-linked mannose residues (45). LAM consists of LM with an arabinan chain attached to the mannan backbone at an unknown position. Every fifth arabinosyl residue is in turn modified with a side chain of approximately seven arabinosyl residues, which can then be further arabinosylated or capped with mannosyl

residues to form ManLAM in slow-growing *Mycobacterium* spp. (57). Specifically, mannose caps are attached through an α(1→5) linkage to the terminal arabinosyl residue, and consist of a single mannose, or dimannosides or trimannosides containing α(1→2) linkages (46). In *Mtb* LAM variably contains a single 5-deoxy-5-methylthio-α-xylofuranose or 5-deoxy-5-methylsulfoxy-α-xylofuranose residue attached to the terminal mannosyl cap residue via an α(1→4) linkage (144, 292, 295). In *M. smegmatis* a small fraction of LAM is capped with inositol phosphate (PILAM) (148), and in *M. chelonae* the arabinosyl termini are not modified (AraLAM) (114).

Peripheral cell wall components are loosely bound to the exterior of the cell envelope. Similar to LAM and PIMs, the noncovalently linked cell envelope lipids may associate or even intercalate with the mycolic acids through hydrophic interactions. These include phthiocerol dimycocerosates (PDIM), the related phenolic glycolipid, trehalose-6,6'-dimycolate (TDM or cord factor), trehalose-6,6'-monomycolate (TMM), sulfolipids, and other acylated trehalose molecules. The "outer layer" described by electron microscopy studies consists primarily of surface polysaccharides (205). These include arabinan, structurally similar to LAM except for loss of the lipid anchor, and glucan, composed of repeating units of α(1→4)-linked glucosyl units (164).

1.3 Pathogenesis

Although *Mtb* is able to initiate disease in multiple organs, 80 to 90% of cases involve the lungs and are termed pulmonary TB (102). Transmission results from inhalation of small aerosol droplets, each typically containing one to ten bacilli and generated by an infected individual through coughing or speaking (229). Once inhaled and deposited in the alveolar space, the bacilli enter macrophages and other resident cells

through receptor-mediated phagocytosis. Infected macrophages release chemokines that recruit activated monocytes and lymphocytes into the lung. The host remodels the site of infection into a cellular mass termed the granuloma, consisting of macrophages, giant cells, T cells, B cells, and fibroblasts (94). In the containment phase a fibrous cutoff composed of collagen and other extracellular matrix components forms the periphery of the granuloma, and serves to physically wall off and effectively sequester the bacterium. In later stages of disease, the fibrous sheath thickens and blood vessel supply decreases. This dense structural architecture is thought to produce a hypoxic environment, as demonstrated through histological studies (293). *Mtb* has clearly evolved strategies to persist in this environment, although the replication and physical state of the bacterium remains under speculation. The cycle of transmission begins again if the immune response of the host weakens, causing caseation and rupture of the granuloma. When this occurs viable, infectious bacilli are released or pushed into the airways resulting in a productive cough (235).

1.3.1 Interactions with the macrophage

Mtb enters the macrophage through receptor-mediated phagocytosis, and this event is believed to occur both upon initial encounter and repeatedly throughout the course of infection. Multiple host cell receptors have been demonstrated to bind various mycobacterial surface molecules. Macrophage receptors include the complement receptors (CR) CR1, CR3, and CR4 (132, 250, 280), the membrane-associated macrophage mannose receptor (MMR) (248), and scavenger receptors (218, 324). Also mediating attachment are the soluble pulmonary surfactant proteins (PSP) –A (211) and PSP-D (89). Specifically, CR3 may bind heparin-binding hemagglutinin

(HBHA/Rv0475) (197) or β-glucan (60), the MMR binds LAM (251) and PIMs (134), PSP-A binds LM, LAM, (262, 263) and the mannosylated 45-kDa protein (Apa/DPEP/ModD/MPT32/Rv1860) (223). Performing a separate function from the macrophage, dendritic cells recognize and take up mycobacteria through the DC-specific intercellular adhesion molecule-3-grabbing nonintegrin (DC-SIGN), a receptor found to bind ManLAM (105, 174). A complete understanding of the phagocyte receptors, uptake mechanisms, and the mycobacterial ligands involved has not yet been reached. However, this is a subject of intense interest because the nature of the receptor-ligand interaction can influence the intensity and type of host cell response and can ultimately decide the fate of the bacterium (249).

Normally phagocytosis leads to fusion of the phagosome with the lysosome, a dynamic process with multiple events termed the endocytic pathway. If formed properly, the combined actions of lysosomal acid hydrolases and the acidic environment (pH 4.5-5.0) of a phagolysosome would normally kill the microorganism (94). However, *Mtb* is able to arrest this maturation at a fairly defined point, as ascertained by several molecular features of the mycobacteria-containing phagosome. These vacuoles retain the GTPase Rab5 but exclude Rab7, molecules required for early and late endosomal interaction events (304). However, mycobacterial phagosomes do acquire the late endosomal/lysosomal marker LAMP-1 (48, 281), suggesting these vacuoles may fuse with other vesicles in a non-traditional, selective manner (234). Mycobacteria-containing phagosomes partially exclude the vacuolar ATPase from other endosomal compartments, explaining the relative lack of acidification (58, 281). Interestingly, the host tryptophan aspartate-containing coat protein (TACO) is retained, providing a mechanistic block for

fusion with lysosomes (90). Defining the mechanism(s) for inhibition of phagosome-lysosome fusion is currently a subject of intense investigation. Inhibition has long been known to be an active process, since the seminal studies of D'Arcy Hart and colleagues demonstrated that live but not dead *Mtb* bacilli could inhibit fusion (14, 120). Recent studies have identified several mycobacterial molecules proposed to interfere with phagosomal maturation, including LAM (99, 301), TDM (139), sulfolipids (112), the serine/threonine kinase PknG/Rv0410c (308), and the acid phosphatase SapM/Rv3310 (302).

A powerful bactericidal mechanism employed by macrophages is the production of reactive oxygen intermediates (ROI) and reactive nitrogen intermediates (RNI). Toxic RNIs are produced through nitric oxide synthase 2 (NOS2)-dependent pathways (172). The primary product of this pathway is nitric oxide which can be converted to other reactive species in a biological setting, yet the RNI responsible for microbial control remains unclear. The role of RNIs in host defense against *Mtb* is well established. Mice with impaired NOS2 pathways are more susceptible to *Mtb* (43, 173, 246), and increased NOS2 activity has been identified in macrophages isolated from TB patients (200, 228). Increased NO levels can even be detected in the exhaled air of TB patients (310), although the protective effect of NO in humans remains to be proven.

The primary ROI is the superoxide anion (O_2-) generated by the NADPH oxidase complex in a reaction known as the "respiratory burst" (98). In contrast to the role of RNIs, the contribution of ROIs in defense against *Mtb* remains controversial. Early experiments with ROI-deficient macrophage cell lines suggested ROIs do not add to the antimycobacterial properties (44). Yet NADPH oxidase gene knockouts do exhibit

modest increases in susceptibility to *Mtb* infection in the mouse model (4, 53). Further supporting a role for ROIs is the finding that chronic granulomatous disease patients with impaired NADPH oxidase systems are more susceptible to TB (158).

Mtb possesses multiple defense strategies to protect against the actions of ROIs and RNIs. ROIs and RNIs can act in synergy, resulting in the particularly destructive peroxynitrite anion formed from the reaction between NO and O_2-. The *Mtb* peroxiredoxin alkylhydroperoxide reductase subunit C (AhpC/Rv2428) has been shown to detoxify this product (37). The *Mtb* gene products of *noxR1* and *noxR3* are able to confer protection against ROI and RNI when expressed in *Salmonella* and *M. smegmatis*, although the mechanisms are unknown (85, 233). The peptide MsrA/Rv0137c is an antioxidant repair enzyme that likely repairs proteins by reducing oxidized methionine residues (83, 273). Recently, a transposon-based mutagenesis screen identified several proteasomal genes in *Mtb* whose loss resulted in hypersensitivity to RNIs. Based on this finding, it was hypothesized that the *Mtb* proteasome serves to degrade or even repair proteins damaged by the action of RNIs (70). More specific to ROI defense, two *Mtb* superoxide dismutases (SOD) are able to convert O_2- into molecular oxygen and H_2O_2. The Mn,Fe-dependent SodA/Rv3846 is a major *Mtb* extracellular protein (322) and is secreted outside of the cell in a unique SecA2-dependent manner (33). The copper-dependent SodC/Rv0432 is a membrane-associated lipoprotein (61) whose activity may protect the bacterial surface against extracellular-generated O_2-. Both SodA and SodC likely contribute to defense against the oxidative burst *in vivo*, as gene knockout studies demonstrate a SodC mutant is readily killed in activated macrophages (219), and a SodA mutant is markedly attenuated in the mouse model (84). The product of the SOD

reaction, H_2O_2, is mildly toxic and is converted to water and molecular oxygen via the sole *Mtb* catalase-peroxidase KatG/Rv1908c, an enzyme also required for activation of isoniazid (166, 176). It should be noted that, in addition to these antioxidant enzymes, *Mtb* LAM and *M. leprae* phenolicglycolipid I (PGL-I) have been shown to scavenge oxygen radicals (41, 42).

1.3.2 The immune response

The immune response mounted against *Mtb* is complex and multifaceted. Bacilli replicate until a critical threshold is reached where the adaptive immune response is initiated, halting replication. Once initiated, the immune response can be described as generally successful in containing, though not eliminating the bacterium. Thus, 90% of infected individuals will not develop clinical disease in their lifetime. However, the remaining 10% will either present with progressive disease within three years of infection, or will develop disease later in life typically when immunity becomes compromised (94). Protective immunity to TB is primarily cellular-based in nature, formally demonstrated through adoptive transfer experiments where splenocytes of BCG- or *Mtb*-sensitized mice conferred protection upon irradiated recipient mice (162, 204). Although a strong antibody response is generated by infected hosts, the prevailing view is that B cells and antibodies do not contribute to protective immunity, a notion that continues to be challenged (108).

The adaptive immune response to *Mtb* infection is thought to be primarily mediated by the CD4+ T cell phenotype. Central to immunity is the recognition of major-histocompatibility complex (MHC) class II-presented antigen by CD4+ T cells that secrete interferon-γ (IFN-γ) to activate the infected macrophages. The key role of IFN-γ

was revealed by the uncontrolled bacterial replication in and early death of IFN-γ knock-out mice (96). Another crucial cytokine is interleukin-12 (IL-12), produced by the macrophage following phagocytosis with the function of driving the Type 1 T cell response and IFN-γ production (94). IFN-γ activates macrophages resulting in induced NOS2 expression and production of ROIs and RNIs in the bacteria-containing phagosomes. The cytokine tumor necrosis factor α (TNF-α) acts in synergy with IFN-γ to further increase macrophage activation. In addition to RNI and ROI production, macrophage activation results in increased intracellular bacterial killing by partially reversing the inhibition of phagosome-lysosome fusion (94). Furthermore, IFN-γ, IL-12 and TNF-α are inflammatory cytokines, and production of these cytokines is necessary for granuloma formation (95).

Recent years have demonstrated an important role for the CD8+ T cell subset in *Mtb* immunity. CD8+ T cells recognize antigen presented by MHC class I molecules. This normally results from processing of antigens by the proteasome in the cytoplasm and loading of the peptides on MHC class I molecules in the endoplasmic reticulum. Interestingly, these activities are contrary to the location of vacuole-localized *Mtb*. Thus, mechanisms by which mycobacterial antigens gain access to the MHC class I presentation pathway have been proposed, including phagosomal permeability (180, 288), cross-presentation of exogenous antigen (137), alternative MHC class I processing pathways (39), or simply *Mtb* escape into the cytosol (298). There are two major effector functions of CD8+ T cells. The first is the production of the cytokines IFN-γ and TNF-α, able to activate macrophages in the same manner as, and augmenting the function of, CD4+ T cells. The second is a cytotoxic function, where CD8+ T cells can lyse target

cells through the actions of perforin and granulysin (277, 278) or through Fas/FasL interactions (278, 296). Recent work suggests the effector functions of CD4+ and CD8+ T cells dominate in the acute phase and persistent phase of infection, respectively (159).

A third T cell subset, the CD1-restricted T cell, is able to recognize *Mtb* nonpeptide antigens in the context of the CD1 molecule (192-195, 264). CD1a-d are antigen-presenting molecules that share homology with MHC class I molecules, yet are distinguished by the ability to bind lipids, glycolipids, and lipoopeptides through a deep hydrophobic binding groove (321). The role of CD1-restricted cells in the immune response to *Mtb* remains to be determined, but both IFN-γ production and cytotoxic activity have been demonstrated for this cell type (94).

The role of innate immunity in TB disease has recently received much attention. *Mtb* interactions with the various macrophage receptors described in section 1.3.1 can influence the host effector response. An example of this is the ManLAM-MR interaction which stimulates antinflammatory cytokines, while AraLAM cannot interact with the MR and behaves like lipopolysaccharide inducing proinflammatory mediators (249). Recently, Toll-like receptor (TLR) proteins have been suggested to play an important regulatory role in the innate immune response against *Mtb* infection. TLRs are pattern recognition receptors (PRRs) expressed on the macrophage and dendritic cell surface that are able to recognize a variety of ligands. TLR2 is the most common receptor involved in *Mtb* recognition, and the list of mycobacterial proteins and glycolipids recognized is continuously expanding, yet the nature of the receptor-ligand interactions remains to be determined (143). TLR engagement results in the regulation of hundreds of host genes including the induction of proinflammatory cytokines. Triggering of the TLR pathway

thus leads to inflammation and granuloma formation; however, the ultimate impact of TLR recognition on the control of *Mtb* infection is not completely understood (34, 143). In contrast to TLRs, nucleotide binding oligomerization domain (NOD)-like receptors (NLRs) are PRRs that recognize bacterial products released into the cytosol (74). The mycobacterial PG component muramyl dipeptide was shown to serve as a ligand for NOD2, and NOD2 activation had a strong synergistic effect on proinflammatory cytokine production induced by TLR2 engagement with the 19-kDa lipoprotein (91). However, the contribution of NLRs in *Mtb* control *in vivo* remains to be established, as NOD2-deficient mice were no more susceptible to *Mtb* infection than wild-type mice (100).

1.4 Immune-based diagnosis

There were an estimated 8.8 million new cases of TB in 2005 (1). Ninety % of these cases occurred in developing countries, where diagnosis relies primarily on identification of acid-fast bacilli (AFB) in sputum samples using conventional light microscopy. However, the sensitivity of this test is estimated to range from 20 to 80 % (276), and is even poorer for diagnosis of pediatric-, extrapulomonary-, and HIV-associated-TB (213, 260). Low sensitivities most often result in delayed diagnosis, enabling disease progression and increasing potential for disease transmission (20). In developed countries, active TB cases are diagnosed with multiple methods, including advanced microscopy methods to detect AFB, and bacteriologic culture paired with nucleic acid amplification tests. Despite these advanced techniques, diagnosis can take several weeks. Developed countries also spend considerable resources on monitoring latent TB infection, identified through skin reactivity to purified protein derivative (PPD), a crude mixture of greater than 200 mycobacterial antigens. However, the PPD test gives

a false-negative reaction in 20 to 30% of TB patients, and fails to distinguish between active disease, prior infection with *Mtb*, BCG vaccination, and exposure to environmental mycobacteria. Thus, for both developing and developed countries alike, a new rapid, sensitive, and specific test is needed to diagnose TB in all of its disease manifestations. The development of such a test has proven to be one of the greatest challenges in TB research. Although molecular biological diagnostic methods display significant potential, especially in identifying drug-resistant TB infections, immune-based methods seem to hold the most promise for a universal diagnostic test and will be the subject of this section.

1.4.1 Identification of diagnostic antigens

Mtb proteins capable of stimulating a T-cell response have been a major research focus due to their potential use in vaccine and diagnostic test development. Although TB disease is effectively controlled through cellular- rather than humoral-based immunity, researchers have maintained an interest in the antibody response elicited against *Mtb*-specific proteins for its potential in serodiagnosis. Thus, the identification of B- and T-cell antigens progressed simultaneously, and the strategies developed for their discovery have coevolved. Initially, single protein antigens were targeted based on 1) their *in-vitro* abundance in the culture filtrate (CF), 2) their ease of biochemical purification, or 3) their immunodominance in animals upon immunization with *in vitro*-grown bacterial preparations (155). These studies resulted in identification of the 14/16-kDa antigen (α-crystallin/Acr/HspX/Rv2031c), the 19-kDa lipoprotein (LpqH/Rv3763), the 38-kDa antigen (antigen5/PstS1/Rv0934), Ag85A (FbpA/P32/Rv3804c), and Ag85B (FbpB/Rv1886c) as serodiagnostic targets (32, 66, 69, 130, 236), and resulted in

identification of the 19-kDa lipoprotein, the Ag85 complex, DnaK/Rv0350, GroES/Rv3418c, the 45-kDa protein, MPT46/TrxC, MPT53/Rv2878c, MPT63/Rv1926c, and the 38-kDa antigen as T-cell antigens (7, 198, 315). These studies also brought about the revelation that B-cell antigens are often potent T-cell antigens, an observation that continues to hold true.

More refined biochemical approaches led to the discovery of additional antigens. Again, the CF protein pool was primarily targeted for analysis based on the hypothesis that secreted proteins are more available to the immune system. Two dimensional (2D) sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) followed by immunoblotting with patients' sera resulted in identification of MPT32, MPT51 (FbpD/Rv3803c), and Ag85c (FbpC/ Rv0129c) as serodiagnostic targets (240). This study was greatly facilitated with preadsorption of the human sera with *E. coli* lysate to deplete cross-reactive antibodies, and this technique was also key in identifying the antigen GlcB (81/88-kDa protein/Mtb81/Rv1837c) (154). Multiple studies developed strategies to analyze protein fractions separated by SDS-PAGE in cellular-based assays (3, 9, 10, 29, 115, 220, 316), notably leading to the discovery of ESAT-6/Rv3875 as a potent T-cell antigen (270). More extensive protein separation strategies were used to identify six novel T-cell antigens by Welding *et al.* (311). Covert *et al.* took an even larger-scale proteomics approach to screen separated cytosolic and CF protein fractions for IFN-γ responses from mouse splenocytes infected with *Mtb* (55). Liquid chromatography (LC) electrospray-ionization (ESI) tandem mass spectrometry (MS/MS) analyses of the reactive fractions resulted in the identification of 17 novel antigens (55).

Strategies for antigen identification have shifted in focus since the sequence of the *Mtb* genome was published (51). *Mtb* is known to alter gene expression patterns to adapt to different environments, and differing protein expression patterns are expected in both the different microenvironments encountered and through the stages of disease progression (181, 214). Therefore, the *in vitro* abundance of *Mtb* proteins may not be relevant to the consortium of proteins expressed and available to the immune system *in vivo*. With this in mind, expression screening approaches were developed, whereby randomly sheared *Mtb* genomic libraries are expressed in *E. coli* and immune responses against all of the ~4000 open reading frames (ORFs) can theoretically be evaluated. Screening these polypeptides with patients' sera or rabbit antisera led to the identification of Mtb81, Mtb48, CFP10/Rv3874, PirG/Rv3810, PTRP, and MtrA/Rv3246c (78, 126, 138, 168, 268). Scientists at Corixa Corp. have continually improved expression strategies to screen against $CD4^+$ T cell responses, resulting in roughly 100 T-cell antigens (5, 78, 79, 269). Additionally, multiple members of the PE/PPE protein families, including PPE55/Rv3347c, PPE18/Mtb39a/Rv1196, and PE-PGRS51/Rv3367, were identified as B- and T-cell antigens with these expression screening strategies (79, 267, 268). These protein families are weakly expressed *in vitro* and have rarely been found in proteomics studies (22), thus highlighting the effectiveness of these identification approaches. The *Mtb* genome sequence also allowed for bioinformatics-based antigen identification approaches, whereby T-cell epitope and signal peptide prediction programs have targeted ORFs for immunological characterization (73, 110).

1.4.2 Cellular based diagnosis

Both $CD4^+$ and $CD8^+$ T-cell subsets are involved in the cellular immune response against *Mtb* infection; however, cellular-based diagnosis has largely focused on evaluating the $CD4^+$ T-cell recall response due to the simpler, more routine methodology as well as the ability of $CD4^+$ to recognize a broader repertoire of antigens. These assays typically measure IFN-γ responses or proliferation of purified peripheral blood mononuclear cells (PBMC) stimulated *in vitro* with antigen. When crude mixtures, such as PPD and culture filtrate proteins (CFP), are used as the antigen very high sensitivities are achieved; however, the specificities are poor. The converse is observed for single, defined protein antigens. Thus, efforts are focused on finding an antigen or antigen combination that yields the best sensitivity. By far the most extensively evaluated T-cell antigens are the ESAT-6 and CFP10 proteins (155). Initial interest in these antigens was primarily due to the absence of their genes in the vaccine strain *M. bovis* BCG compared to *Mtb* (21, 118). Indeed, these antigens have been reported to be highly sensitive and specific in distinguishing TB patients from noninfected individuals, as well as *Mtb*-infected from BCG-vaccinated and *M. avium*-complex-infected individuals (13, 163, 299). Furthermore, a report suggests ESAT-6 recognition by healthy household contacts correlates well with development of active TB, an attractive attribute of a diagnostic test (82). The ESAT-6 and CFP-10 antigens are the basis for the recently developed, commercially available T-SPOT.*TB* and QuantiFERON®-TB Gold tests (156).

Cellular-based diagnosis studies have also focused on antigens that can elicit a delayed-type hypersensitivity (DTH) response, in hopes of improving on the current PPD skin test. Again, antigens restricted to *Mtb* not found in BCG were targeted for

evaluation. CFP-10, ESAT-6, and MPT64/Rv1980c could distinguish *Mtb*-infected from BCG- and *M. avium*- infected guinea pigs (50, 86). However, the performance of CFP-10 and ESAT-6 in human clinical trials remains to be validated, while MPT64 produced disappointing results in humans (155, 312) . An especially promising DTH antigen is DPPD/Rv3663c. A protein limited to the MTBC, this antigen has provided superior responses to PPD in both guinea pigs (52) and humans (38). Furthermore, DPPD provides a clearer distinction between a positive and negative skin reaction than PPD, and in a large-scale study produced a positive reaction in a percentage of healthy, infected individuals that was close to the estimated prevalence (38). However, the ability of DPPD to distinguish from BCG vaccination has not been determined.

1.4.3 Serodiagnosis

Although a clear role for antibody-mediated immunity in protection against *Mtb* infection remains to be established, exposure to *Mtb* elicits production of antibodies that can be used as markers of TB infection. Efforts to develop a diagnostic test based on antibody-detection for active TB have been underway for decades, and dozens of such tests are currently marketed in developing countries (155, 275). However, a recent systematic review of the performance of these commercial tests concluded that none function well enough to replace sputum smear microscopy (275). Most of these commercial tests are based on the 38-kDa antigen, an antigen that has been touted as the "gold standard" because it gave the best sensitivity among the single candidate antigens identified in initial studies (155). However, studies have since demonstrated that 38-kDa antigen sensitivities vary among geographic populations (2), and that antibodies against this antigen are primarily present in patients with advanced stages of disease (32, 240,

241), thereby greatly limiting its diagnostic utility and providing an explanation for the poor performance of today's commercial tests.

The most challenging aspect of developing a TB serodiagnostic test is identifying the optimal set of antigenic components to be used in the assay. Crude mycobacterial preparations, such as PPD, whole-cell lysate (WCL), and CFP, share antigenic components with other bacterial genera such as *Nocardia* and *Corynebacterium* (67), and have been shown lack to specificity for *Mtb* infection. Several glycolipids, including LAM, TDM, sulfolipid-I, diacyltrehalose (TAT), and triacyltrehalose (TAT), are recognized by a significant percentage of TB patients' sera, but also result in poor specificities (145, 290). In contrast, the plethora of TB serological studies generally agree that single, purified native or recombinant proteins provide excellent specificities (155). A review of these studies suggests there are greater than 50 protein antigens that provide a sensitivity of at least ten %, and reactivity summaries for well-characterized antigen candidates are reviewed in references (40) and (155). However, difficulties arise in reviewing the serological properties of even a single antigen, as numerous studies often provide conflicting sensitivity estimates. For instance, there are greater than 15 estimates for the 38-kDa antigen recognition among pulmonary TB patients, with sensitivities ranging from 21 to 100% (2). Although some variation can be expected across geographic populations, disparities also arise from the methods used to identify healthy controls and to classify TB patient cohorts, the quality of the sera, and the level of scientific rigor used in the analyses. Fittingly, guidelines have recently been suggested by The Standards for Reporting of Diagnostic Accuracy initiative for publishing results related to infectious disease diagnosis (19). Instead of reviewing the sensitivity values of

each antigen identified to date, it is more appropriate to review the research that will lead to a successful serodiagnostic assay, namely addressing the criteria for an ideal antigen.

Laal *et al.* performed a series of systematic analyses to directly address the hypothesis that *Mtb* antigen profiles differ depending on patient disease status (239, 240). Specifically, sera samples were used to probe 2D maps of CF proteins. Of the >100 visible protein spots, only 24 to 26 protein species were recognized by sera from patients with advanced, cavitary TB. Interestingly, a subset of proteins (12 of 26) were recognized by sera from all classes of TB patients, including noncavitary, smear positive, smear negative, non-HIV-infected (HIV-TB+), or HIV-coinfected (HIV+TB+). Thus, there was remarkable homogeneity in the antigen profiles recognized among the different groups, although the patient sample size was limited. Furthermore, specific antigens like MPT51 and GlcB were recognized by all TB patient groups tested, while antigens like the 38-kDa and MPT32 were associated with the later stages of disease progression. Recently, other laboratories have also made associations between the serodominance of specific antigens and a particular TB disease state. Patients with inactive (past) TB possessed higher antibody titers against HspX/Rv2031c, FdxA/Rv2007c, and ESAT-6, while active TB patients' sera produced more dominant responses against the 38-kDa antigen, Ald/Rv2780, and Rv2626 (72). The finding that the antigen HspX is associated with latent TB has been independently validated by other groups (30, 31, 76) and is reflective of its functional role during nonreplicating persistence (317, 318) and its transcriptional profile in the mouse model (258).

As the HIV epidemic spreads to areas of high TB incidence, there is an increasing need for specific tests to diagnose HIV+TB+ individuals. Serodiagnosis would be

especially attractive in such cases, as HIV infection is associated with a decrease in the levels of $CD4^+$ T cells, limiting the utility of cellular-based assays. Thus, recent efforts have identified antigens recognized by retrospective HIV+TB+ patients' sera months before progression to clinical TB. Laal *et al.* have shown the antigen GlcB to be recognized by 75% of such serum samples (153). Furthermore, 19 of 21 (90%) retrospective serum samples recognized a combination of GlcB and MPT51 (265). In addition, a separate report demonstrated that Sulfolipid IV was significantly recognized by HIV+TB+ retrospective sera (177). Considering the high rate of latent TB reactivation and the accelerated rate of TB disease progression associated with HIV-coinfection, these individuals could greatly benefit from preventative therapy if identified in a timely manner.

Close contacts of infectious TB patients have a 30% chance of becoming infected, thus a diagnostic test is especially needed in low-prevalence countries for surveillance purposes (27, 152). The current PPD skin test is not able to distinguish BCG-vaccinated individuals from latent TB infection, and gives a substantial number of false-negative reactions. Unfortunately, the promising T-cell antigenicity of the region of deletion 1 (RD1) 1 antigens (described in section 1.4.2) is not matched with strong B-cell antigenicity. When six proteins encoded by the RD1 region were characterized, 16% or less of TB patients possessed antibodies to any one protein (36). However, continued analyses of the missing genes in BCG versus *Mtb* may lead to serodiagnostic candidates. Recently, the isocitrate dehydrogenase enzymes Icd1/Rv3339c and Icd2/Rv0066c, present in both *Mtb* and BCG, were claimed to distinguish BCG-vaccinated healthy

individuals from TB patients (18). The reasons for this differential recognition remain elusive.

The repertoire of antigens recognized by the immune system may be expected to differ depending on the infecting *Mtb* strain. However, whole-cell lysate protein profiles between H37Rv and the clinical isolate CDC1551 were found to be surprisingly similar (24), given the number of predicted differences in their gene products (93). Furthermore, a separate study specifically searched for serodiagnostic antigens and compared the CF protein profiles of *Mtb* strains H37Rv, CDC1551, and a Korean TB isolate, and found only a minority of proteins were differentially expressed (17). The ubiquitous presence of dominant B cell antigens is encouraging, as a universal diagnostic test should be able to detect infection with the numerous strains of *Mtb* found throughout the world.

Strategies will need to be addressed for producing reagents for serodiagnostic assays. Recombinant production of *Mtb* proteins in *E. coli* has become a relatively routine and facile procedure, and ultimately would be fairly inexpensive for large-scale antigen production. However, a body of evidence suggests these recombinant antigens exhibit reduced immunoreactivity compared to their native counterparts. Specifically, recombinant Ag85c and MPT32 produced in *E. coli* were recognized by less than half of the patients' sera that recognized the native forms, suggesting the recombinant forms lacked antigenic determinants (241). These determinants may be specific posttranslational modifications (PTMs) and/or conformational epitopes not reproduced upon protein processing in *E. coli*. Indeed, the serological importance of PTMs has been demonstrated. Deglycosylation of MPT32 was shown to result in reduced seroreactivity (157), and the methylation of HBHA significantly contributed to serological recognition

(259). Furthermore, heterologous hosts designed to mimic *Mtb* protein production have recently been investigated. *M. smegmatis* was able produce a seroreactive methylated HBHA molecule (259), and the yeast *Pichia pastoris* was able to produce a more seroreactive CFP32 molecule compared to the *E. coli* recombinant form (23).

Despite the extensive efforts in *Mtb* antigen characterization, the best sensitivities achieved to date for any antigen or combination of antigens is 75 to 80% for smear-positive TB patients (152). The reasons for the lack of serum antibodies in 20% of patients are purely speculative. Some TB patients may lack antibody production due to an especially robust Th1 response. Alternatively, the formation of antigen-antibody complexes may prevent detection. In support of the latter hypothesis, Singh *et al.* (266) found antimycobacterial antibodies could be detected in the urine when absent from the corresponding patient's sputum samples, suggesting low antibody titers trapped in immune complexes may lead to lowered antigen recognition. However, this scenario is actually encouraging, because detection strategies can always be improved if an antibody response is generated at some level.

The combined research suggests that a successful TB serodiagnostic assay will ultimately be based on a cocktail of antigens. With this in mind, the poly-protein TbF6 was developed by genetically fusing together four known antigenic sequences. The resulting recombinant protein provided a higher sensitivity than any of the individual antigen components (138). A fairly crude immunochromatographic assay was developed, termed "MAPIA", which was based on a cocktail of ten different recombinant *Mtb* antigens. MAPIA was shown to be superior to the enzyme-linked immunoadsorbant

assay (ELISA), because the total sensitivity achieved was equal to sum of its components (171).

1.5 Posttranslationally-modified mycobacterial proteins

Recent years have allowed for a deeper understanding of the function of PTMs in mycobacteria: the great diversity and sheer prevalence of mycobacteria PTMs is startling. The functional roles of PTMs in the basic biology of *Mtb* are beginning to be elucidated, though the roles in pathogenesis remain enigmatic.

1.5.1 Glycoproteins

1.5.1.1 Identification

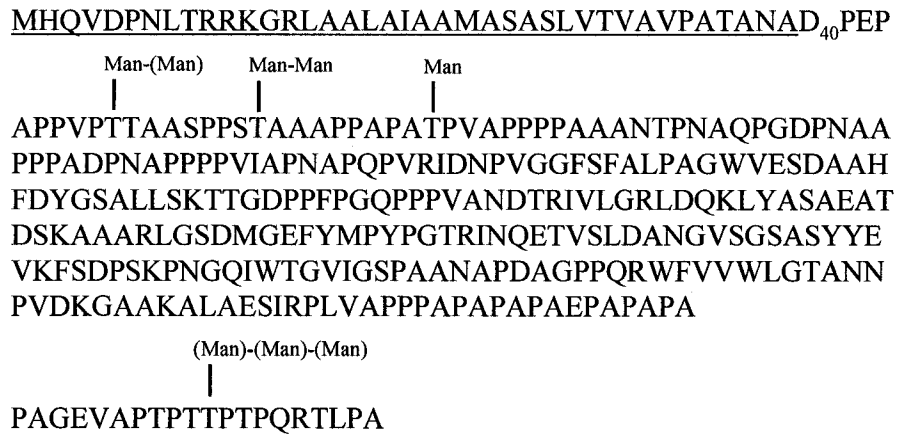
Glycosylation of proteins was originally thought to be a unique phenomenon of higher eukaryotes. This assumption is understandable, as glycoproteins are either absent or very scarce in the most commonly studied bacterial species including *E. coli*, *Salmonella* spp, and *Bacillus subtilis* (187). It was not until 1976, with the identification of a surface layer (S-layer) glycoprotein of *Halobacterium salinarium* (186), that the presence of prokaryotic glycoproteins began to gain widespread acceptance. Multiple discoveries of glycosylated proteins within the domain *Archaea* and then *Bacteria* soon followed (187), and at least 70 bacterial proteins have been reported thus far (297). First evidence for mycobacterial glycoproteins resulted from analyses of the previously identified the 38-kDa antigen (66), suggesting this antigen was intimately associated with or covalently linked with sugars similar to arabinomannan (65, 68). Two studies soon followed that employed the lectin Concanavalin A (ConA) and further identified potentially glycosylated forms of the 38-kDa protein and the 45-kDa protein from *Mtb* (88), and in *M. bovis* detected glycoforms of the 19-kDa lipoprotein homologue, the 38-

kDa antigen (PstS1/Mb0959) homologue, and MPB70/Mb2900 (92) later determined to in fact be MPB83/Mb2898 (129). The ability of *M. smegmatis* to glycosylate the heterologously-expressed 19-kDa lipoprotein was demonstrated (103), and a molecular approach further localized the glycosylated region to two threonine (Thr) clusters (128). Expanding on this work, the same group fused peptide sequences from eleven different *Mtb* lipoproteins to the 19-kDa lipoprotein signal peptide, and demonstrated eight of the sequences (originating from the 19-kDa lipoprotein, the 38-kDa protein, MPT83/Rv2873, GlnH/Rv0411c, LppN/Rv2270, LppQ/Rv2431, LprI/Rv1541c, and SodC) produced ConA-reactive products when expressed in *M. smegmatis* (127). These results also demonstrated that the neural network NetOglyc, a glycoprotein prediction method trained on eukaryotic O-glycosylation sites (117), could reliably identify mycobacterial glycoproteins (127). Furthermore, glycosylation of lipoproteins appears to be common in mycobacteria, as recombinant LprA/Rv1270c and LprG/Rv1411c were recently shown to be ConA-reactive (212), unpublished observations. A proteomic study analyzed *Mtb* secreted proteins with a top down MS approach, and a fraction enriched by ConA chromatography contained a 9-kDa protein modified with up to seven hexose units and two 20-kDa proteins each modified with up to 20 hexose units; however the protein identities were not provided (104). A recent study employed a commercial glycan detection kit and reported that PPE68/Rv3873, a protein of the RD1 locus and a member of the PE/PPE protein family, is glycosylated near the C-terminus (71). Although these results are preliminary, glycosylation of members of the large PE/PPE family is intriguing considering their contributions to virulence (165, 224, 269).

1.5.1.2 Structural determination

Despite the relative abundance of mycobacterial glycoproteins, detailed structural information has been provided for only two proteins, first for the 45/47-kDa protein (80, 81), and later for MPB83 from *M. bovis* (188). These studies were also the first to provide definitive chemical proof for protein glycosylation, as they were not limited to ConA-based analysis, a technique hindered by the recognition of contaminating glycosylated molecules in protein preparations. Dobos *et al.* combined multiple MS-based methods with N-terminal amino acid sequencing to identify and characterize four glycosylated Thr residues in the 45/47-kDa protein (80, 81). These residues were O-linked with α-D-mannose, α-D-mannobiose, or α-D-mannotriose, and each mannose chain was determined to contain only α(1→2) linkages (80, 81) (Fig. 1.1). A subsequent report demonstrated heterogeneity in the level of MPT32 glycosylation, variably possessing zero to ten but dominantly seven mannose residues (232). Michell *et al.* combined genetic and biochemical methods to identify an O-glycosylated Thr-Thr doublet near the N-terminus of the *M. bovis* secreted antigen MPB83 (188) (Fig. 1.1). The modifying sugar(s) were demonstrated to be predominantly mannose and mannobiose, with low levels of mannotriose. In contrast to MPT32, the mannobiose was found to be of an α(1→3) linkage (188).

MPT32[1]



MPB83[2]

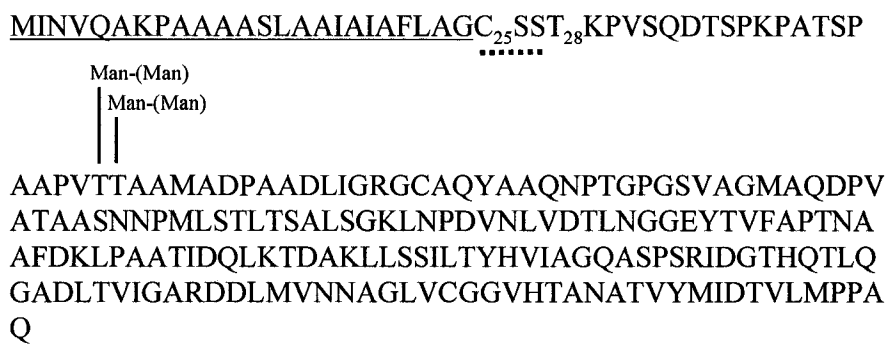


Fig. 1.1 MPT32 and MPB83 mannosylation sites. The regions underlined with a solid line denote the signal peptides. Parentheses denote variable levels of glycosylation. [1]Aspartic acid residue 40 (D_{40}) is the experimentally determined N-terminal amino acid for the major CF product. Mannose linkages are of the $\alpha(1\rightarrow2)$ configuration. [2]The truncated sequence underlined with the dashed line includes the putative acylated Cys residue 25 (C_{25}), preceding the experimentally determined N-terminal Thr residue 28 (T_{28}). Mannose linkages are of the $\alpha(1\rightarrow3)$ configuration. *Man*, mannose

1.5.1.3 Biosynthesis and function

Aspects of *Mtb* glycoprotein biosynthesis have recently begun to be elucidated. Mycobacterial glycoproteins are modified in regions rich with Thr, Ser, Ala, and Pro residues. Substitution of these residues flanking glycosylation sites inhibited glycosylation, suggesting these sequences may be required for protein-mannosyltransferase (PMT) substrate recognition (54, 128). Using membrane extracts from *M. smegmatis*, Cooper *et al.* employed a cell-free assay to glycosylate MPT32 peptide sequences with radiolabelled GDP-mannose, and demonstrated the membrane-associated PMT activity was sensitive to amphomycin, suggesting lipid-linked C35/C50-polyprenol-Man intermediates are required for biosynthesis (54). Most recently Vanderven *et al.* employed a similar assay to establish PMT activity for the protein Rv1002c, a membrane protein found to be dependent on the conserved PMT Asp-Glu motif party responsible for catalytic activity (300). Furthermore, this work demonstrated that, similar to eukaryotic organisms, sec-dependent protein translocation is required for *Mtb* mannoprotein biosynthesis, and proposed Rv1002c may physically associate with the Sec complex (300). Although Rv1002c catalyzes the initial mannosylation step, additional PMTs may be required for mannan chain elongation. The presence of multiple PMTs could also explain the observed differences in MPT32 and MPB83 oligomannoside linkages.

Protein glycosylation serves numerous roles in biological processes of eukaryotes, including protein stability, protein localization, protein folding, receptor recognition, and enzymatic activity (125, 169). Accordingly, defects in protein glycosylation have been linked to multiple mammalian diseases (169). The functional roles of prokaryotic protein

glycosylation are less well understood. However, emerging lines of evidence suggest glycosylation contributes significantly to pathogenesis, serving such functions as host cell surface recognition and adhesion, maintenance of bacterial shape, protection against proteolytic digestion, immune invasion, enzyme-substrate binding, and direct interaction with components of the immune system (252, 297). Single, isolated studies have addressed glycoprotein function in mycobacteria. In terms of the immune response, MPT32 glycosylation was required for both T-cell and antibody recognition, suggesting glycosylated sequences often compose immunodominant epitopes (135, 157, 232). Furthermore, MPT32 was recently suggested to be a potential adhesion, capable of binding the C-type lectin PSP-A found on phagocytic cells (223). A more physiological role for glycosylation was proposed for the 19-kDa lipoprotein by Herrmann *et al.* In an attempt to identify glycosylation sites, suspected O-glycosylated Thr residues were substituted with Val residues, and the recombinant forms were produced in *M. smegmatis* (128). In addition to a reduction in ConA-binding activity, the Thr to Val substitutions also resulted in proteolytic cleavage at the N-terminus, suggesting glycosylation may serve to protect against proteolysis (128).

1.5.2 Lipoproteins

1.5.2.1 Structure and biosynthesis

In order to direct the localization of specific proteins to membrane structures, bacteria modify proteins by acylating the N-terminal Cys residue. Specifically, the Cys residue universally possesses a diacylglycerol unit linked to the thiol group via a thioether bond, and an additional acyl group is variably linked through an amide bond to the N-terminus (285). In bacteria membrane phopholipids serve as donors for these acyl

groups (47, 141). Biosynthesis of bacterial lipoproteins depends on the presence of a Type II signal peptide containing the consensus sequence [LVI][ASTVI][GAS]C, commonly referred to as the "lipobox" (16, 305). Following sec-dependent translocation across the cytoplasmic membrane the diacylglyceride unit is transferred by the prolipoprotein diacylglycerryl transferase (Lgt) (244), thought to be the Rv1614 gene product in *Mtb* (51). The signal peptidase (SPase) II (Lsp) then cleaves signal peptide resulting in the new mature N-terminus beginning with the acylated Cys (314). The *Mtb* LspA/Rv1539 enzyme has been partially characterized, and an *lspA* deletion mutant was dispensable for *in vitro* growth yet was severely attenuated in the mouse model, suggesting the effects of LspA activity, although likely pleiotropic, contribute to virulence (243). The third enzyme in lipoprotein biosynthesis is the lipoprotein N-acyl transferase (Lnt), an enzyme present in Gram-negative but not Gram-positive bacteria. The existence of this enzyme in *Mtb* remains to be determined, as the originally annotated Lnt (Rv2051c) (51) was recently shown to function as a polyprenol monophosphomannose synthase, termed Ppm1 (116). However, Ppm1 may be an example of a "moonlighting" protein, an increasingly common phenomenon found in bacteria (142). Finally, although *Mtb* protein lipidation is presumed to be of the same arrangement as conserved in all bacteria, structural analyses of the acylated Cys for an *Mtb* lipoprotein is still lacking.

1.5.2.2 Identification and localization

Recently, 99 putative lipoproteins were identified by searching the *Mtb* genome with a modified lipobox sequence pattern developed specifically for Gram-positive bacteria (283, 284). The functions of these lipoproteins are very diverse. Many of these

lipoproteins share significant homologies with fully secreted, non-lipidated *Mtb* proteins, suggesting acylation serves primarily to direct compartmentalization of these functions (284). Very few *Mtb* lipoproteins have been experimentally characterized, and direct chemical or structural evidence for an acylated Cys residue in *Mtb* is still lacking. Evidence for *Mtb* protein acylation has largely been provided for solely by the presence in the detergent layer following Triton X-114 phase partitioning; however, this method may not discriminate between true lipoproteins and membrane proteins. A single study by Young *et al.* metabolically labeled lipoproteins by supplying bacteria exogenous radiolabelled fatty acid, and several proteins, including the 19-kDa and 38-kDa proteins, were observed following two-dimensional (2-D) gel electrophoresis and autoradiography (315). Radiolabelling experiments in *E. coli* also suggested the proteins SodC and Rv0679c to be acylated, although lipidation by this host may be irrelevant to *Mtb* (61, 179). Future lipoprotein identification may be aided by analyses of the LspA deletion mutant (243), or through the use of globomycin, an inhibitor of LspA recently employed to characterize two lipoproteins from *M. smegmatis* (106). Recently, LppX/Rv2945c from *Mtb* was confirmed as a lipoprotein through radiolabelling, and was shown to participate in the translocation of phthiocerol dimycocerosate (DIM) to the surface of *Mtb* (282). This provides the intriguing possibility that this large unique family of annotated lipoproteins with unknown function may be involved in the transport of various molecules across the complex mycobacterial cell envelope.

In Gram-positive bacteria, acylation serves to anchor the lipoprotein to the outer face of the plasma membrane by means of hydrophobic interaction. Furthermore, the lipoproteins of Gram-positive bacteria are thought to be the functional equivalents of

periplasmic proteins in Gram-negative bacteria (201, 285). Yet lipoprotein localization may differ significantly when considering the more complex nature of the mycobacterial cell wall. Structural parallels can be drawn between the mycobacterial and Gram-negative cell envelopes, and lipoproteins may in fact associate with the outer lipid layer composed predominantly of mycolic acids. Transport mechanisms to this layer have yet to be defined, but may be analogous to the *E. coli* Lol system (199). Accordingly, flow cytometry and electron microscopy techniques have identified multiple lipoproteins associated with the periphery of the cell, including SodC (313), MPB83 (119, 307), the 38-kDa protein (PstS1) (87), and the LppX homologue Mb2970c from *M. bovis* (161). However, secreted forms of the latter three with N-terminal truncations have also been identified in the CF (8, 161, 188). Thus mycobacteria may share similarities with *Bacillus subtilis*, where recent proposals of Antelmann *et al.* suggest CF forms of lipoproteins result from 1) "shedding" – passive release of acylated lipoproteins, and/or 2) "shaving" – proteolytic release of lipoproteins from the cell wall (11, 291).

1.5.3 Phosphoproteins

Through conserved mechanisms proteins are phosphorylated and dephosphorylated in order to transmit signals intracellularly and to regulate protein function. In the simplest model of the ubiquitous bacterial two-component regulatory system, a membrane-bound sensor histidine kinase (HK) responds to an environmental stimulus causing a conformational change that allows autophosporylation of a His residue within the cytosolic domain. The phosphate group is then transferred to the Asp residue of the response regulator (RR), and the phosphorylated RR directs gene expression or protein function (279). The *Mtb* genome encodes for 11 HK-RR pairs and several orphan

HKs and RRs (51), easily identified because of their highly conserved domains; however, only a subset of these have been experimentally verified. Recombinant forms of the HK-RR pairs MprB/Rv0982-MprA/Rv0981 (319), TrxcR-TrxcS (123), SenX3/Rv0490-RegX3/Rv0491 (131), DevS/DosS/Rv3132c-DevR/DosR/Rv3133c (237), and Rv3220c-PdtaR (196) were demonstrated to autophosphorylate and then transfer phosphate, as expected, and the RR MtrA/Rv3246c was shown to accept phosphate from the *E. coli* HK CheA (303). Recent years have identified functional roles for some of these modules. The SenX3-RegX3 module responds to phosphate starvation (109), the RR MtrA modulates *Mtb* proliferation in macrophages (97), the PhoR/Rv0758-PhoP/Rv0757 module like responds to intracellular pH (178) and regulates biosynthesis of complex lipids (111, 170, 309), and the MprB-MprA module responds to stress and regulates expression of the alternative sigma factors SigB/Rv2710 and SigE/rv1221 (124) as well as the α-crystallin gene *acr2/rv2051c* (207). The most well-characterized RR DevR responds to hypoxic conditions resulting in a specific phenotypic response (209, 257), and was found to bind a DNA consensus sequence present upstream of hypoxic genes in a phosphorylation-state-dependent manner (209). DevR phosphorylation is mediated by two HKs, DevS and DosT (231), and both bind CO, O_2, and NO through heme-dependent manners (140, 271). The active CO- and NO- bound forms and the inactive O2-bound forms provide an explanation for the nearly identical gene expression patterns produced when *Mtb* is exposed to NO and hypoxic conditions (306).

The presence of eleven putative Ser/Thr kinases is thought to compensate for the comparatively low number of two-component system modules in *Mtb* (51). Similar to the two-component system, initial studies verified many of the annotated Ser/Thr kinases

through *in vitro* phosphorylation experiments, and these are reviewed in references (22) and (15). Recent studies have focused on exploring the functional consequences of the Ser/Thr kinases and identifying their phosphorylated substrates. PknA/Rv0015c and PknB/Rv0014c were recently implicated in regulating cell shape, possibly affecting septation and cell division (146). PknF/Rv1746 phosphorylates Rv1747 (191), an ABC transporter possibly involved in sugar transport (77). Analyses of a *pknG/rv0410c* knockout strain displayed an attenuated growth phenotype in culture, attributed to the role of PknG in sensing and regulating glutamine levels (56). Contrary to these roles, PknG was shown to be secreted into macrophages and was reported to inhibit phagosome-lysosome fusion, most likely through phosphorylation of host substrates (308). EmbR/rv3252c, a transcriptional regulator of arabinosyltransferases, is phosphorylated by PknH/Rv1266c (190); however, EmbR was also subsequently shown to be phosphorylated by PknA and PknB (256). Furthermore, PknH was able to phosphorylate multiple substrates, including Rv0681 and DacB1/Rv3330 (323). These combined observations suggest there may be a complex interplay between the Ser/Thr kinases and their substrates, making single gene knockout studies difficult to interpret. The more stable nature of Ser- and Thr- phosphoesters in comparison to phospho-His and phosphor-Asp requires a phosphatase enzyme (261). The sole Ser/Thr phosphatase in *Mtb* is PstP/Rv0018c (51). Although the crystal structure of PstP has been reported (222), the functional consequences of *pstP* deletion have yet to be determined. Although the *Mtb* genome does not contain any recognizable tyrosine (Tyr) kinase homologues, two Tyr phosphatase enzymes (PtpA/Rv2234 and PtpB/Rv0153c) are present (150); however, secretion of these enzymes suggest they may purely play a role in virulence by

interfering with host signaling pathways (151). *Mtb* has been estimated to contain on the order of 500-1,000 phosphorylated Ser/Thr kinase protein substrates (113), and future studies are needed to address the phosphoproteome. Recent reports of phosphorylated forms of the FASII enzymes MtFabH/Rv0533c, KasA/Rv2245, and KasB/Rv2246 (25, 189), the Pks13/Rv3800c condensase involved in the last step of mycolic acid biosynthesis (25, 221), and the Fe-dependent superoxide dismutase SodA (12) suggest phosphorylation of enzymes is a common mechanism of enzyme regulation in mycobacteria.

1.5.4 Atypical post-translationally modified proteins

A clear functional role of a PTM is in the case of adenylylation. The *Mtb* genome encodes four homologues of the glutamine synthetase (GS) enzyme (*glnA1-glnA4*), responsible for the ATP-dependent incorporation of ammonia into glutamate to form glutamine; however, only GlnA1/Rv2220 can be detected, isolated, and assayed in growing cultures and is the only homologue considered essential for *in vitro* growth (122, 294). GlnA1 may be important for pathogenesis, as pathogenic but not nonpathogenic mycobacteria export large quantities of this enzyme (225). Furthermore, a *glnA1* mutant had a reduced ability to multiply within macrophages and was avirulent in the guinea pig model (294), and this attenuated strain has been explored as a potential vaccine (160). Using an MS-based approach, Tyr residue 406 of GlnA1 was shown to be modified with an adenylyl group (183). Like other bacterial GS enzymes, GlnA1 functions as a dodecameric oligomer (121), and in other bacteria, GS activity is regulated by adenylylation of the monomeric subunits, with the number of adenylylated subunits inversely proportional to GS catalytic activity (254, 255). The putative adenylyl-

transferase GlnE/Rv2221c, an enzyme essential for *in vitro* growth, is thought to be the enzyme responsible for this modification (208). When available pools of ammonia quickly decrease, GlnA1 adenylylation by GlnE is thought be a faster method of enzyme regulation as opposed to transcriptional control of *glnA1* (210). In *E. coli* the adenylylation activity of GlnE is itself regulated by uridylation from a separate enzyme; however, the presence of this regulatory cascade and uridylation in mycobacteria remains to be demonstrated (226).

The *Mtb* HBHA has been shown to bind sulfated glycoconjugates present on nonphagocytic cells (Menozzi 1996), and has been postulated to be important for extrapulmonary dissemination (214). Recombinant HBHA produced in *M. smegmatis* was found to be of a greater molecular mass than recombinant HBHA from *E. coli* (217). This difference was originally attributed to glucosylation (217); however, a subsequent analysis found this mass difference was instead due to modification with a complex pattern of up to 26 methyl groups on 13 lysine (Lys) residues, each variably mono- or dimethylated (216). All methylated Lys residues are within a series of Lys-rich repeats in the C-terminal domain, and these surface-exposed repeats have been shown to be responsible for the binding activity (185, 215). Methylation protects this region from proteolytic degradation, as methylation conferred stability upon HBHA when exposed to trypsin and murine alveolar lavage fluids (216). HBHA methylation also clearly influences antigenicity, as methylated but not non-methylated HBHA forms were 1) recognized infected TB patients' sera (259, 320), 2) required for strong T cell responses from infected individuals (289), and 3) provided protection against *Mtb* in the mouse model (289). Yet methylation is neither specific to pathogenic mycobacteria nor limited

to HBHA, as *M. smegmatis* possesses both a methylated HBHA homologue (26) and a suspected methylated laminan-binding protein (LBP) (217). The methyltransferase enzyme remains to be identified, but S-adenosylmethionine-dependent HBHA methyltransferase activity has been identified in cell-wall-derived mycobacterial extracts (216). The HBHA N-terminal domain is required for *in vivo* but not *in vitro* methylation (75, 136), and HBHA is exposed on the surface of mycobacteria (185). These combined observations suggest export may be required for methylation; however the N-terminal domain lacks a recognizable sec-dependent signal peptide (184). Identification of the HBHA methylation mechanism may help in the design of novel strategies to prevent TB extrapulomary dissemination, thought by some to be crucial for TB reactivation (167).

Forms of ESAT-6 (6 kDa early secreted antigen target), a protein currently under intense investigation, possess an N-terminal acetylated Thr residue, and the acetylation was found to influence physical interactions with ESAT-6 and the co-transcribed CFP10 (203). In fact, a recent re-evaluation of archived MS data suggests N-acetylation of proteins may be a common phenomenon in mycobacteria as opposed to other prokaryotes (230). Whether acetylation also modulates other *Mtb* protein-protein interactions remains to be determined.

Two isolated reports of unconventional PTMs exist for *Mycobacterium* spp. Evidence for protein ADP-ribosylation was provided by labeling various mycobacterial cell extracts with [α-^{32}P]-NAD. Specifically, two proteins from *M smegmatis* were found to be enzymatically ADP-ribosylated, and this modification was suspected to be on cysteine (Cys) and asparagine residues. (253) Furthermore, an ADP-ribosyltransferase homologue has been annotated in *Mycobacterium avium* (206). However, the role of

ADP-ribosylation in mycobacteria remains unknown. A recent proteome-wide study identified S-nitrosylated Cys residues on 29 proteins following treatment of *Mtb* cultures with mildly acidified nitrite (227). These conditions were designed to mimic conditions found in the phagosome of activated macrophages, where a major postulated mycobactericidal activity of reactive nitrogen intermediates (RNI) is enzyme inactivation through S-nitrosylation (227). The *Mtb* S-nitroso proteome was heavily enriched in both genes predicted essential for *in vitro* growth (245), and genes shown to be important for *Mtb* virulence and persistence (227). Furthermore, the enzymatic activities of two proteins, a lipoamide dehydrogenase (LpD/Rv0462) and a mycobacterial proteasome ATPase (Mpa/Rv1604), were reduced as a result of S-nitrosylation (227). The classification of S-nitrosylation as a true PTM is debatable, as S-nitrosylation results from an exogenous chemical modification rather than from a mycobacterial enzymatic process. However, this same study identified three endogenously biotinylated proteins, and identified two of these species as subunits of acetyl CoA carboxylase (AccA3/Rv3285) and pyruvate carboxylase (Pca/Rv2967c) (227). Both of these enzymes are conserved and biotinylated in most organisms (242).

1.6 Literature cited

1. *Global tuberculosis control - surveillance, planning, financing*, in *WHO Report 2007*. 2007, World Health Organization: Geneva.

2. **Abebe, F., C. Holm-Hansen, H. G. Wiker, and G. Bjune**. 2007. Progress in serodiagnosis of *Mycobacterium tuberculosis* infection. Scand J Immunol **66**(2-3): 176-91.

3. **Abou-Zeid, C., E. Filley, J. Steele, and G. A. Rook**. 1987. A simple new method for using antigens separated by polyacrylamide gel electrophoresis to stimulate lymphocytes in vitro after converting bands cut from Western blots into antigen-bearing particles. J Immunol Methods **98**(1): 5-10.

4. **Adams, L. B., M. C. Dinauer, D. E. Morgenstern, and J. L. Krahenbuhl.** 1997. Comparison of the roles of reactive oxygen and nitrogen intermediates in the host response to *Mycobacterium tuberculosis* using transgenic mice. Tuber Lung Dis **78**(5-6): 237-46.

5. **Alderson, M. R., T. Bement, C. H. Day, L. Zhu, D. Molesh, Y. A. Skeiky, R. Coler, D. M. Lewinsohn, S. G. Reed, and D. C. Dillon.** 2000. Expression cloning of an immunodominant family of *Mycobacterium tuberculosis* antigens using human CD4(+) T cells. J Exp Med **191**(3): 551-60.

6. **Allen, B. W.** 1969. *Mycobacterium tuberculosis* strain H37Rv. J Med Lab Technol **26**(4): 389-90.

7. **Andersen, A. B. and E. B. Hansen.** 1989. Structure and mapping of antigenic domains of protein antigen b, a 38,000-molecular-weight protein of *Mycobacterium tuberculosis*. Infect Immun **57**(8): 2481-8.

8. **Andersen, A. B., L. Ljungqvist, and M. Olsen.** 1990. Evidence that protein antigen b of *Mycobacterium tuberculosis* is involved in phosphate metabolism. J Gen Microbiol **136**(3): 477-80.

9. **Andersen, P., A. B. Andersen, A. L. Sorensen, and S. Nagai.** 1995. Recall of long-lived immunity to *Mycobacterium tuberculosis* infection in mice. J Immunol **154**(7): 3359-72.

10. **Andersen, P. and I. Heron.** 1993. Simultaneous electroelution of whole SDS-polyacrylamide gels for the direct cellular analysis of complex protein mixtures. J Immunol Methods **161**(1): 29-39.

11. **Antelmann, H., H. Tjalsma, B. Voigt, S. Ohlmeier, S. Bron, J. M. van Dijl, and M. Hecker.** 2001. A proteomic view on genome-based signal peptide predictions. Genome Res **11**(9): 1484-502.

12. **Archambaud, C., M. A. Nahori, J. Pizarro-Cerda, P. Cossart, and O. Dussurget.** 2006. Control of *Listeria* superoxide dismutase by phosphorylation. J Biol Chem **281**(42): 31812-22.

13. **Arend, S. M., P. Andersen, K. E. van Meijgaarden, R. L. Skjot, Y. W. Subronto, J. T. van Dissel, and T. H. Ottenhoff.** 2000. Detection of active tuberculosis infection by T cell responses to early-secreted antigenic target 6-kDa protein and culture filtrate protein 10. J Infect Dis **181**(5): 1850-4.

14. **Armstrong, J. A. and P. D. Hart.** 1971. Response of cultured macrophages to *Mycobacterium tuberculosis*, with observations on fusion of lysosomes with phagosomes. J Exp Med **134**(3 Pt 1): 713-40.

15. **Av-Gay, Y. and M. Everett.** 2000. The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. Trends Microbiol **8**(5): 238-44.

16. **Babu, M. M., M. L. Priya, A. T. Selvan, M. Madera, J. Gough, L. Aravind, and K. Sankaran**. 2006. A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. J Bacteriol **188**(8): 2761-73.

17. **Bahk, Y. Y., S. A. Kim, J. S. Kim, H. J. Euh, G. H. Bai, S. N. Cho, and Y. S. Kim**. 2004. Antigens secreted from *Mycobacterium tuberculosis*: identification by proteomics approach and test for diagnostic marker. Proteomics **4**(11): 3299-307.

18. **Banerjee, S., A. Nandyala, R. Podili, V. M. Katoch, K. J. Murthy, and S. E. Hasnain**. 2004. *Mycobacterium tuberculosis* (Mtb) isocitrate dehydrogenases show strong B cell response and distinguish vaccinated controls from TB patients. Proc Natl Acad Sci U S A **101**(34): 12652-7.

19. **Banoo, S., D. Bell, P. Bossuyt, A. Herring, D. Mabey, F. Poole, P. G. Smith, N. Sriram, C. Wongsrichanalai, R. Linke, R. O'Brien, M. Perkins, J. Cunningham, P. Matsoso, C. M. Nathanson, P. Olliaro, R. W. Peeling, and A. Ramsay**. 2006. Evaluation of diagnostic tests for infectious diseases: general principles. Nat Rev Microbiol **4**(12 Suppl): S20-32.

20. **Behr, M. A., S. A. Warren, H. Salamon, P. C. Hopewell, A. Ponce de Leon, C. L. Daley, and P. M. Small**. 1999. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. Lancet **353**(9151): 444-9.

21. **Behr, M. A., M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small**. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science **284**(5419): 1520-3.

22. **Belisle, J. T., M. Braunstein, I. Rosenkrands, and P. Andersen**, The Proteome of *Mycobacterium tuberculosis*, in Tuberculosis and the Tubercle Bacillus, S.T. Cole, Editor. 2005, ASM Press: Washington, D.C. p. 235-60.

23. **Benabdesselem, C., D. M. Fathallah, R. C. Huard, H. Zhu, M. A. Jarboui, K. Dellagi, J. L. Ho, and R. M. Barbouche**. 2006. Enhanced patient serum immunoreactivity to recombinant *Mycobacterium tuberculosis* CFP32 produced in the yeast Pichia pastoris compared to *Escherichia coli* and its potential for serodiagnosis of tuberculosis. J Clin Microbiol **44**(9): 3086-93.

24. **Betts, J. C., P. Dodson, S. Quan, A. P. Lewis, P. J. Thomas, K. Duncan, and R. A. McAdam**. 2000. Comparison of the proteome of *Mycobacterium tuberculosis* strain H37Rv with clinical isolate CDC 1551. Microbiology **146 Pt 12**: 3205-16.

25. **Bhatt, A., V. Molle, G. S. Besra, W. R. Jacobs, Jr., and L. Kremer**. 2007. The *Mycobacterium tuberculosis* FAS-II condensing enzymes: their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. Mol Microbiol **64**(6): 1442-54.

26. **Biet, F., M. Angela de Melo Marques, M. Grayon, E. K. Xavier da Silveira, P. J. Brennan, H. Drobecq, D. Raze, M. C. Vidal Pessolani, C. Locht, and F. D. Menozzi**. 2007. *Mycobacterium smegmatis* produces an HBHA homologue which is not involved in epithelial adherence. Microbes Infect **9**(2): 175-82.

27. **Binkin, N. J., A. A. Vernon, P. M. Simone, E. McCray, B. I. Miller, C. W. Schieffelbein, and K. G. Castro**. 1999. Tuberculosis prevention and control activities in the United States: an overview of the organization of tuberculosis services. Int J Tuberc Lung Dis **3**(8): 663-74.

28. **Boddinghaus, B., T. Rogall, T. Flohr, H. Blocker, and E. C. Bottger**. 1990. Detection and identification of mycobacteria by amplification of rRNA. J Clin Microbiol **28**(8): 1751-9.

29. **Boesen, H., B. N. Jensen, T. Wilcke, and P. Andersen**. 1995. Human T-cell responses to secreted antigen fractions of *Mycobacterium tuberculosis*. Infect Immun **63**(4): 1491-7.

30. **Bothamley, G. H.** 2004. Epitope-specific antibody levels demonstrate recognition of new epitopes and changes in titer but not affinity during treatment of tuberculosis. Clin Diagn Lab Immunol **11**(5): 942-51.

31. **Bothamley, G. H., J. S. Beck, R. C. Potts, J. M. Grange, T. Kardjito, and J. Ivanyi**. 1992. Specificity of antibodies and tuberculin response after occupational exposure to tuberculosis. J Infect Dis **166**(1): 182-6.

32. **Bothamley, G. H., R. Rudd, F. Festenstein, and J. Ivanyi**. 1992. Clinical value of the measurement of *Mycobacterium tuberculosis* specific antibody in pulmonary tuberculosis. Thorax **47**(4): 270-5.

33. **Braunstein, M., B. J. Espinosa, J. Chan, J. T. Belisle, and W. R. Jacobs, Jr.** 2003. SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. Mol Microbiol **48**(2): 453-64.

34. **Brightbill, H. D., D. H. Libraty, S. R. Krutzik, R. B. Yang, J. T. Belisle, J. R. Bleharski, M. Maitland, M. V. Norgard, S. E. Plevy, S. T. Smale, P. J. Brennan, B. R. Bloom, P. J. Godowski, and R. L. Modlin**. 1999. Host defense mechanisms triggered by microbial lipoproteins through toll-like receptors. Science **285**(5428): 732-6.

35. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole**. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc Natl Acad Sci U S A **99**(6): 3684-9.

36. **Brusasca, P. N., R. Colangeli, K. P. Lyashchenko, X. Zhao, M. Vogelstein, J. S. Spencer, D. N. McMurray, and M. L. Gennaro**. 2001. Immunological

characterization of antigens encoded by the RD1 region of the *Mycobacterium tuberculosis* genome. Scand J Immunol **54**(5): 448-52.

37. **Bryk, R., P. Griffin, and C. Nathan**. 2000. Peroxynitrite reductase activity of bacterial peroxiredoxins. Nature **407**(6801): 211-5.

38. **Campos-Neto, A., V. Rodrigues-Junior, D. B. Pedral-Sampaio, E. M. Netto, P. J. Ovendale, R. N. Coler, Y. A. Skeiky, R. Badaro, and S. G. Reed**. 2001. Evaluation of DPPD, a single recombinant *Mycobacterium tuberculosis* protein as an alternative antigen for the Mantoux test. Tuberculosis (Edinb) **81**(5-6): 353-8.

39. **Canaday, D. H., C. Ziebold, E. H. Noss, K. A. Chervenak, C. V. Harding, and W. H. Boom**. 1999. Activation of human CD8+ alpha beta TCR+ cells by *Mycobacterium tuberculosis* via an alternate class I MHC antigen-processing pathway. J Immunol **162**(1): 372-9.

40. **Chan, E. D., L. Heifets, and M. D. Iseman**. 2000. Immunologic diagnosis of tuberculosis: a review. Tuber Lung Dis **80**(3): 131-40.

41. **Chan, J., X. D. Fan, S. W. Hunter, P. J. Brennan, and B. R. Bloom**. 1991. Lipoarabinomannan, a possible virulence factor involved in persistence of *Mycobacterium tuberculosis* within macrophages. Infect Immun **59**(5): 1755-61.

42. **Chan, J., T. Fujiwara, P. Brennan, M. McNeil, S. J. Turco, J. C. Sibille, M. Snapper, P. Aisen, and B. R. Bloom**. 1989. Microbial glycolipids: possible virulence factors that scavenge oxygen radicals. Proc Natl Acad Sci U S A **86**(7): 2453-7.

43. **Chan, J., K. Tanaka, D. Carroll, J. Flynn, and B. R. Bloom**. 1995. Effects of nitric oxide synthase inhibitors on murine infection with *Mycobacterium tuberculosis*. Infect Immun **63**(2): 736-40.

44. **Chan, J., Y. Xing, R. S. Magliozzo, and B. R. Bloom**. 1992. Killing of virulent *Mycobacterium tuberculosis* by reactive nitrogen intermediates produced by activated murine macrophages. J Exp Med **175**(4): 1111-22.

45. **Chatterjee, D., S. W. Hunter, M. McNeil, and P. J. Brennan**. 1992. Lipoarabinomannan. Multiglycosylated form of the mycobacterial mannosylphosphatidylinositols. J Biol Chem **267**(9): 6228-33.

46. **Chatterjee, D., K. Lowell, B. Rivoire, M. R. McNeil, and P. J. Brennan**. 1992. Lipoarabinomannan of *Mycobacterium tuberculosis*. Capping with mannosyl residues in some strains. J Biol Chem **267**(9): 6234-9.

47. **Chattopadhyay, P. K. and H. C. Wu**. 1977. Biosynthesis of the covalently linked diglyceride in murein lipoprotein of *Escherichia coli*. Proc Natl Acad Sci U S A **74**(12): 5318-22.

48. **Clemens, D. L. and M. A. Horwitz**. 1995. Characterization of the *Mycobacterium tuberculosis* phagosome and evidence that phagosomal maturation is inhibited. J Exp Med **181**(1): 257-70.

49. **Cohen, J.** 2006. Infectious disease. Extensively drug-resistant TB gets foothold in South Africa. Science **313**(5793): 1554.

50. **Colangeli, R., J. S. Spencer, P. Bifani, A. Williams, K. Lyashchenko, M. A. Keen, P. J. Hill, J. Belisle, and M. L. Gennaro**. 2000. MTSA-10, the product of the Rv3874 gene of *Mycobacterium tuberculosis*, elicits tuberculosis-specific, delayed-type hypersensitivity in guinea pigs. Infect Immun **68**(2): 990-3.

51. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell**. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393**(6685): 537-44.

52. **Coler, R. N., Y. A. Skeiky, P. J. Ovendale, T. S. Vedvick, L. Gervassi, J. Guderian, S. Jen, S. G. Reed, and A. Campos-Neto**. 2000. Cloning of a *Mycobacterium tuberculosis* gene encoding a purifed protein derivative protein that elicits strong tuberculosis-specific delayed-type hypersensitivity. J Infect Dis **182**(1): 224-33.

53. **Cooper, A. M., B. H. Segal, A. A. Frank, S. M. Holland, and I. M. Orme**. 2000. Transient loss of resistance to pulmonary tuberculosis in p47(phox-/-) mice. Infect Immun **68**(3): 1231-4.

54. **Cooper, H. N., S. S. Gurcha, J. Nigou, P. J. Brennan, J. T. Belisle, G. S. Besra, and D. Young**. 2002. Characterization of mycobacterial protein glycosyltransferase activity using synthetic peptide acceptors in a cell-free assay. Glycobiology **12**(7): 427-34.

55. **Covert, B. A., J. S. Spencer, I. M. Orme, and J. T. Belisle**. 2001. The application of proteomics in defining the T cell antigens of *Mycobacterium tuberculosis*. Proteomics **1**(4): 574-86.

56. **Cowley, S., M. Ko, N. Pick, R. Chow, K. J. Downing, B. G. Gordhan, J. C. Betts, V. Mizrahi, D. A. Smith, R. W. Stokes, and Y. Av-Gay**. 2004. The *Mycobacterium tuberculosis* protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo. Mol Microbiol **52**(6): 1691-702.

57. **Crick, D. C., P. J. Brennan, and M. R. McNeil**, The Cell Wall of Myobacterium Tuberculosis, in Tuberculosis, S. Garay and W. Rom, Editors. 2004, Lippincott, Williams & Wilkins: Philadelphia. p. 115-34.

58. **Crowle, A. J., R. Dahl, E. Ross, and M. H. May**. 1991. Evidence that vesicles containing living, virulent *Mycobacterium tuberculosis* or *Mycobacterium avium* in cultured human macrophages are not acidic. Infect Immun **59**(5): 1823-31.

59. **Crubezy, E., B. Ludes, J. D. Poveda, J. Clayton, B. Crouau-Roy, and D. Montagnon**. 1998. Identification of Mycobacterium DNA in an Egyptian Pott's disease of 5,400 years old. C R Acad Sci III **321**(11): 941-51.

60. **Cywes, C., H. C. Hoppe, M. Daffe, and M. R. Ehlers**. 1997. Nonopsonic binding of *Mycobacterium tuberculosis* to complement receptor type 3 is mediated by capsular polysaccharides and is strain dependent. Infect Immun **65**(10): 4258-66.

61. **D'Orazio, M., S. Folcarelli, F. Mariani, V. Colizzi, G. Rotilio, and A. Battistoni**. 2001. Lipid modification of the Cu,Zn superoxide dismutase from *Mycobacterium tuberculosis*. Biochem J **359**(Pt 1): 17-22.

62. **Daffe, M., P. J. Brennan, and M. McNeil**. 1990. Predominant structural features of the cell wall arabinogalactan of *Mycobacterium tuberculosis* as revealed through characterization of oligoglycosyl alditol fragments by gas chromatography/mass spectrometry and by 1H and 13C NMR analyses. J Biol Chem **265**(12): 6734-43.

63. **Daffe, M. and P. Draper**. 1998. The envelope layers of mycobacteria with reference to their pathogenicity. Adv Microb Physiol **39**: 131-203.

64. **Daffe, M. and G. Etienne**. 1999. The capsule of *Mycobacterium tuberculosis* and its implications for pathogenicity. Tuber Lung Dis **79**(3): 153-69.

65. **Daniel, T. M.** 1989. The chemical composition of immunoaffinity-purified *Mycobacterium tuberculosis* antigen 5. Am Rev Respir Dis **139**(6): 1566-7.

66. **Daniel, T. M. and P. A. Anderson**. 1978. The isolation by immunoabsorbent affinity chromatography and physicochemical characterization of *Mycobacterium tuberculosis* antigen 5. Am Rev Respir Dis **117**(3): 533-9.

67. **Daniel, T. M. and S. M. Debanne**. 1987. The serodiagnosis of tuberculosis and other mycobacterial diseases by enzyme-linked immunosorbent assay. Am Rev Respir Dis **135**(5): 1137-51.

68. **Daniel, T. M., N. J. Gonchoroff, J. A. Katzmann, and G. R. Olds**. 1984. Specificity of *Mycobacterium tuberculosis* antigen 5 determined with mouse monoclonal antibodies. Infect Immun **45**(1): 52-5.

69. **Daniel, T. M. and B. W. Janicki**. 1978. Mycobacterial antigens: a review of their isolation, chemistry, and immunological properties. Microbiol Rev **42**(1): 84-113.

70. **Darwin, K. H., S. Ehrt, J. C. Gutierrez-Ramos, N. Weich, and C. F. Nathan**. 2003. The proteasome of *Mycobacterium tuberculosis* is required for resistance to nitric oxide. Science **302**(5652): 1963-6.

71. **Daugelat, S., J. Kowall, J. Mattow, D. Bumann, R. Winter, R. Hurwitz, and S. H. Kaufmann**. 2003. The RD1 proteins of *Mycobacterium tuberculosis*: expression in *Mycobacterium smegmatis* and biochemical characterization. Microbes Infect **5**(12): 1082-95.

72. **Davidow, A., G. V. Kanaujia, L. Shi, J. Kaviar, X. Guo, N. Sung, G. Kaplan, D. Menzies, and M. L. Gennaro**. 2005. Antibody profiles characteristic of *Mycobacterium tuberculosis* infection state. Infect Immun **73**(10): 6846-51.

73. **De Groot, A. S., A. Bosma, N. Chinai, J. Frost, B. M. Jesdale, M. A. Gonzalez, W. Martin, and C. Saint-Aubin**. 2001. From genome to vaccine: in silico predictions, ex vivo verification. Vaccine **19**(31): 4385-95.

74. **Delbridge, L. M. and M. X. O'Riordan**. 2007. Innate recognition of intracellular bacteria. Curr Opin Immunol **19**(1): 10-6.

75. **Delogu, G., A. Bua, C. Pusceddu, M. Parra, G. Fadda, M. J. Brennan, and S. Zanetti**. 2004. Expression and purification of recombinant methylated HBHA in *Mycobacterium smegmatis*. FEMS Microbiol Lett **239**(1): 33-9.

76. **Demissie, A., E. M. Leyten, M. Abebe, L. Wassie, A. Aseffa, G. Abate, H. Fletcher, P. Owiafe, P. C. Hill, R. Brookes, G. Rook, A. Zumla, S. M. Arend, M. Klein, T. H. Ottenhoff, P. Andersen, and T. M. Doherty**. 2006. Recognition of stage-specific mycobacterial antigens differentiates between acute and latent infections with *Mycobacterium tuberculosis*. Clin Vaccine Immunol **13**(2): 179-86.

77. **Deol, P., R. Vohra, A. K. Saini, A. Singh, H. Chandra, P. Chopra, T. K. Das, A. K. Tyagi, and Y. Singh**. 2005. Role of *Mycobacterium tuberculosis* Ser/Thr kinase PknF: implications in glucose transport and cell division. J Bacteriol **187**(10): 3415-20.

78. **Dillon, D. C., M. R. Alderson, C. H. Day, T. Bement, A. Campos-Neto, Y. A. Skeiky, T. Vedvick, R. Badaro, S. G. Reed, and R. Houghton**. 2000. Molecular and immunological characterization of *Mycobacterium tuberculosis* CFP-10, an immunodiagnostic antigen missing in *Mycobacterium bovis* BCG. J Clin Microbiol **38**(9): 3285-90.

79. **Dillon, D. C., M. R. Alderson, C. H. Day, D. M. Lewinsohn, R. Coler, T. Bement, A. Campos-Neto, Y. A. Skeiky, I. M. Orme, A. Roberts, S. Steen, W.**

Dalemans, R. Badaro, and S. G. Reed. 1999. Molecular characterization and human T-cell responses to a member of a novel *Mycobacterium tuberculosis* mtb39 gene family. Infect Immun **67**(6): 2941-50.

80. **Dobos, K. M., K. H. Khoo, K. M. Swiderek, P. J. Brennan, and J. T. Belisle**. 1996. Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. J Bacteriol **178**(9): 2498-506.

81. **Dobos, K. M., K. Swiderek, K. H. Khoo, P. J. Brennan, and J. T. Belisle**. 1995. Evidence for glycosylation sites on the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. Infect Immun **63**(8): 2846-53.

82. **Doherty, T. M., A. Demissie, J. Olobo, D. Wolday, S. Britton, T. Eguale, P. Ravn, and P. Andersen**. 2002. Immune responses to the *Mycobacterium tuberculosis*-specific antigen ESAT-6 signal subclinical infection among contacts of tuberculosis patients. J Clin Microbiol **40**(2): 704-6.

83. **Douglas, T., D. S. Daniel, B. K. Parida, C. Jagannath, and S. Dhandayuthapani**. 2004. Methionine sulfoxide reductase A (MsrA) deficiency affects the survival of *Mycobacterium smegmatis* within macrophages. J Bacteriol **186**(11): 3590-8.

84. **Edwards, K. M., M. H. Cynamon, R. K. Voladri, C. C. Hager, M. S. DeStefano, K. T. Tham, D. L. Lakey, M. R. Bochan, and D. S. Kernodle**. 2001. Iron-cofactored superoxide dismutase inhibits host responses to *Mycobacterium tuberculosis*. Am J Respir Crit Care Med **164**(12): 2213-9.

85. **Ehrt, S., M. U. Shiloh, J. Ruan, M. Choi, S. Gunzburg, C. Nathan, Q. Xie, and L. W. Riley**. 1997. A novel antioxidant gene from *Mycobacterium tuberculosis*. J Exp Med **186**(11): 1885-96.

86. **Elhay, M. J., T. Oettinger, and P. Andersen**. 1998. Delayed-type hypersensitivity responses to ESAT-6 and MPT64 from *Mycobacterium tuberculosis* in the guinea pig. Infect Immun **66**(7): 3454-6.

87. **Espitia, C., M. Elinos, R. Hernandez-Pando, and R. Mancilla**. 1992. Phosphate starvation enhances expression of the immunodominant 38-kilodalton protein antigen of *Mycobacterium tuberculosis*: demonstration by immunogold electron microscopy. Infect Immun **60**(7): 2998-3001.

88. **Espitia, C. and R. Mancilla**. 1989. Identification, isolation and partial characterization of *Mycobacterium tuberculosis* glycoprotein antigens. Clin Exp Immunol **77**(3): 378-83.

89. **Ferguson, J. S., D. R. Voelker, F. X. McCormack, and L. S. Schlesinger**. 1999. Surfactant protein D binds to *Mycobacterium tuberculosis* bacilli and lipoarabinomannan via carbohydrate-lectin interactions resulting in reduced phagocytosis of the bacteria by macrophages. J Immunol **163**(1): 312-21.

90. **Ferrari, G., H. Langen, M. Naito, and J. Pieters**. 1999. A coat protein on phagosomes involved in the intracellular survival of mycobacteria. Cell **97**(4): 435-47.

91. **Ferwerda, G., S. E. Girardin, B. J. Kullberg, L. Le Bourhis, D. J. de Jong, D. M. Langenberg, R. van Crevel, G. J. Adema, T. H. Ottenhoff, J. W. Van der Meer, and M. G. Netea**. 2005. NOD2 and toll-like receptors are nonredundant recognition systems of *Mycobacterium tuberculosis*. PLoS Pathog **1**(3): 279-85.

92. **Fifis, T., C. Costopoulos, A. J. Radford, A. Bacic, and P. R. Wood**. 1991. Purification and characterization of major antigens from a *Mycobacterium bovis* culture filtrate. Infect Immun **59**(3): 800-7.

93. **Fleischmann, R. D., D. Alland, J. A. Eisen, L. Carpenter, O. White, J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J. F. Kolonay, W. C. Nelson, L. A. Umayam, M. Ermolaeva, S. L. Salzberg, A. Delcher, T. Utterback, J. Weidman, H. Khouri, J. Gill, A. Mikula, W. Bishai, W. R. Jacobs Jr, Jr., J. C. Venter, and C. M. Fraser**. 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J Bacteriol **184**(19): 5479-90.

94. **Flynn, J. L. and J. Chan**. 2001. Immunology of tuberculosis. Annu Rev Immunol **19**: 93-129.

95. **Flynn, J. L. and J. Chan**. 2005. What's good for the host is good for the bug. Trends Microbiol **13**(3): 98-102.

96. **Flynn, J. L., J. Chan, K. J. Triebold, D. K. Dalton, T. A. Stewart, and B. R. Bloom**. 1993. An essential role for interferon gamma in resistance to *Mycobacterium tuberculosis* infection. J Exp Med **178**(6): 2249-54.

97. **Fol, M., A. Chauhan, N. K. Nair, E. Maloney, M. Moomey, C. Jagannath, M. V. Madiraju, and M. Rajagopalan**. 2006. Modulation of *Mycobacterium tuberculosis* proliferation by MtrA, an essential two-component response regulator. Mol Microbiol **60**(3): 643-57.

98. **Forman, H. J. and M. Torres**. 2002. Reactive oxygen species and cell signaling: respiratory burst in macrophage signaling. Am J Respir Crit Care Med **166**(12 Pt 2): S4-8.

99. **Fratti, R. A., J. Chua, I. Vergne, and V. Deretic**. 2003. *Mycobacterium tuberculosis* glycosylated phosphatidylinositol causes phagosome maturation arrest. Proc Natl Acad Sci U S A **100**(9): 5437-42.

100. **Gandotra, S., S. Jang, P. J. Murray, P. Salgame, and S. Ehrt**. 2007. Nucleotide-Binding Oligomerization Domain Protein 2-Deficient Mice Control Infection with *Mycobacterium tuberculosis*. Infect Immun **75**(11): 5127-34.

101. **Gandy, M.**, Life without Germs: Contested Episodes in the History of Tuberculosis, in The Return of the White Plague, M.G.a.A. Zumla, Editor. 2003, Verso: London. p. 15-38.

102. **Garay, S.**, Pulmonary Tuberculosis, in Tuberculosis, S. Garay and W.N. Rom, Editors. 2004, Lippincott Williams & Wilkins: Philadelphia. p. 345-94.

103. **Garbe, T., D. Harris, M. Vordermeier, R. Lathigra, J. Ivanyi, and D. Young**. 1993. Expression of the *Mycobacterium tuberculosis* 19-kilodalton antigen in *Mycobacterium smegmatis*: immunological analysis and evidence of glycosylation. Infect Immun **61**(1): 260-7.

104. **Ge, Y., M. El-Naggar, S. K. Sze, H. B. Oh, T. P. Begley, F. W. McLafferty, H. Boshoff, and C. E. Barry, 3rd**. 2003. Top down characterization of secreted proteins from *Mycobacterium tuberculosis* by electron capture dissociation mass spectrometry. J Am Soc Mass Spectrom **14**(3): 253-61.

105. **Geijtenbeek, T. B., S. J. Van Vliet, E. A. Koppel, M. Sanchez-Hernandez, C. M. Vandenbroucke-Grauls, B. Appelmelk, and Y. Van Kooyk**. 2003. Mycobacteria target DC-SIGN to suppress dendritic cell function. J Exp Med **197**(1): 7-17.

106. **Gibbons, H. S., F. Wolschendorf, M. Abshire, M. Niederweis, and M. Braunstein**. 2007. Identification of two *Mycobacterium smegmatis* lipoproteins exported by a SecA2-dependent pathway. J Bacteriol **189**(14): 5090-100.

107. **Gilleron, M., V. F. Quesniaux, and G. Puzo**. 2003. Acylation state of the phosphatidylinositol hexamannosides from *Mycobacterium bovis* bacillus Calmette Guerin and *Mycobacterium tuberculosis* H37Rv and its implication in Toll-like receptor response. J Biol Chem **278**(32): 29880-9.

108. **Glatman-Freedman, A.** 2006. The role of antibody-mediated immunity in defense against *Mycobacterium tuberculosis*: advances toward a novel vaccine strategy. Tuberculosis (Edinb) **86**(3-4): 191-7.

109. **Glover, R. T., J. Kriakov, S. J. Garforth, A. D. Baughn, and W. R. Jacobs, Jr.** 2007. The two-component regulatory system senX3-regX3 regulates phosphate-dependent gene expression in *Mycobacterium smegmatis*. J Bacteriol **189**(15): 5495-503.

110. **Gomez, M., S. Johnson, and M. L. Gennaro**. 2000. Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. Infect Immun **68**(4): 2323-7.

111. **Gonzalo Asensio, J., C. Maia, N. L. Ferrer, N. Barilone, F. Laval, C. Y. Soto, N. Winter, M. Daffe, B. Gicquel, C. Martin, and M. Jackson**. 2006. The virulence-associated two-component PhoP-PhoR system controls the biosynthesis

of polyketide-derived lipids in *Mycobacterium tuberculosis*. J Biol Chem **281**(3): 1313-6.

112. **Goren, M. B., P. D'Arcy Hart, M. R. Young, and J. A. Armstrong**. 1976. Prevention of phagosome-lysosome fusion in cultured macrophages by sulfatides of *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A **73**(7): 2510-4.

113. **Greenstein, A. E., C. Grundner, N. Echols, L. M. Gay, T. N. Lombana, C. A. Miecskowski, K. E. Pullen, P. Y. Sung, and T. Alber**. 2005. Structure/function studies of Ser/Thr and Tyr protein phosphorylation in *Mycobacterium tuberculosis*. J Mol Microbiol Biotechnol **9**(3-4): 167-81.

114. **Guerardel, Y., E. Maes, E. Elass, Y. Leroy, P. Timmerman, G. S. Besra, C. Locht, G. Strecker, and L. Kremer**. 2002. Structural study of lipomannan and lipoarabinomannan from Mycobacterium chelonae. Presence of unusual components with alpha 1,3-mannopyranose side chains. J Biol Chem **277**(34): 30635-48.

115. **Gulle, H., B. Schoel, and S. H. Kaufmann**. 1990. Direct blotting with viable cells of protein mixtures separated by two-dimensional gel electrophoresis. J Immunol Methods **133**(2): 253-61.

116. **Gurcha, S. S., A. R. Baulard, L. Kremer, C. Locht, D. B. Moody, W. Muhlecker, C. E. Costello, D. C. Crick, P. J. Brennan, and G. S. Besra**. 2002. Ppm1, a novel polyprenol monophosphomannose synthase from *Mycobacterium tuberculosis*. Biochem J **365**(Pt 2): 441-50.

117. **Hansen, J. E., O. Lund, N. Tolstrup, A. A. Gooley, K. L. Williams, and S. Brunak**. 1998. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconj J **15**(2): 115-30.

118. **Harboe, M., T. Oettinger, H. G. Wiker, I. Rosenkrands, and P. Andersen**. 1996. Evidence for occurrence of the ESAT-6 protein in *Mycobacterium tuberculosis* and virulent *Mycobacterium bovis* and for its absence in *Mycobacterium bovis* BCG. Infect Immun **64**(1): 16-22.

119. **Harboe, M., H. G. Wiker, G. Ulvund, B. Lund-Pedersen, A. B. Andersen, R. G. Hewinson, and S. Nagai**. 1998. MPB70 and MPB83 as indicators of protein localization in mycobacterial cells. Infect Immun **66**(1): 289-96.

120. **Hart, P. D., J. A. Armstrong, C. A. Brown, and P. Draper**. 1972. Ultrastructural study of the behavior of macrophages toward parasitic mycobacteria. Infect Immun **5**(5): 803-7.

121. **Harth, G., D. L. Clemens, and M. A. Horwitz**. 1994. Glutamine synthetase of *Mycobacterium tuberculosis*: extracellular release and characterization of its enzymatic activity. Proc Natl Acad Sci U S A **91**(20): 9342-6.

122. **Harth, G., S. Maslesa-Galic, M. V. Tullius, and M. A. Horwitz**. 2005. All four *Mycobacterium tuberculosis* glnA genes encode glutamine synthetase activities but only GlnA1 is abundantly expressed and essential for bacterial homeostasis. Mol Microbiol **58**(4): 1157-72.

123. **Haydel, S. E., N. E. Dunlap, and W. H. Benjamin, Jr.** 1999. In vitro evidence of two-component system phosphorylation between the *Mycobacterium tuberculosis* TrcR/TrcS proteins. Microb Pathog **26**(4): 195-206.

124. **He, H., R. Hovey, J. Kane, V. Singh, and T. C. Zahrt**. 2006. MprAB is a stress-responsive two-component system that directly regulates expression of sigma factors SigB and SigE in *Mycobacterium tuberculosis*. J Bacteriol **188**(6): 2134-43.

125. **Helenius, A. and M. Aebi**. 2001. Intracellular functions of N-linked glycans. Science **291**(5512): 2364-9.

126. **Hendrickson, R. C., J. F. Douglass, L. D. Reynolds, P. D. McNeill, D. Carter, S. G. Reed, and R. L. Houghton**. 2000. Mass spectrometric identification of mtb81, a novel serological marker for tuberculosis. J Clin Microbiol **38**(6): 2354-61.

127. **Herrmann, J. L., R. Delahay, A. Gallagher, B. Robertson, and D. Young**. 2000. Analysis of post-translational modification of mycobacterial proteins using a cassette expression system. FEBS Lett **473**(3): 358-62.

128. **Herrmann, J. L., P. O'Gaora, A. Gallagher, J. E. Thole, and D. B. Young**. 1996. Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from *Mycobacterium tuberculosis*. Embo J **15**(14): 3547-54.

129. **Hewinson, R. G., S. L. Michell, W. P. Russell, R. A. McAdam, and W. R. Jacobs, Jr.** 1996. Molecular characterization of MPT83: a seroreactive antigen of *Mycobacterium tuberculosis* with homology to MPT70. Scand J Immunol **43**(5): 490-9.

130. **Hewitt, J., A. R. Coates, D. A. Mitchison, and J. Ivanyi**. 1982. The use of murine monoclonal antibodies without purification of antigen in the serodiagnosis of tuberculosis. J Immunol Methods **55**(2): 205-11.

131. **Himpens, S., C. Locht, and P. Supply**. 2000. Molecular characterization of the mycobacterial SenX3-RegX3 two-component system: evidence for autoregulation. Microbiology **146 Pt 12**: 3091-8.

132. **Hirsch, C. S., J. J. Ellner, D. G. Russell, and E. A. Rich**. 1994. Complement receptor-mediated uptake and tumor necrosis factor-alpha-mediated growth inhibition of *Mycobacterium tuberculosis* by human alveolar macrophages. J Immunol **152**(2): 743-53.

133. **Holtz, T. H. and J. P. Cegielski**. 2007. Origin of the term XDR-TB. Eur Respir J **30**(2): 396.

134. **Hoppe, H. C., B. J. de Wet, C. Cywes, M. Daffe, and M. R. Ehlers**. 1997. Identification of phosphatidylinositol mannoside as a mycobacterial adhesin mediating both direct and opsonic binding to nonphagocytic mammalian cells. Infect Immun **65**(9): 3896-905.

135. **Horn, C., A. Namane, P. Pescher, M. Riviere, F. Romain, G. Puzo, O. Barzu, and G. Marchal**. 1999. Decreased capacity of recombinant 45/47-kDa molecules (Apa) of *Mycobacterium tuberculosis* to stimulate T lymphocyte responses related to changes in their mannosylation pattern. J Biol Chem **274**(45): 32023-30.

136. **Host, H., H. Drobecq, C. Locht, and F. D. Menozzi**. 2007. Enzymatic methylation of the *Mycobacterium tuberculosis* heparin-binding haemagglutinin. FEMS Microbiol Lett **268**(2): 144-50.

137. **Houde, M., S. Bertholet, E. Gagnon, S. Brunet, G. Goyette, A. Laplante, M. F. Princiotta, P. Thibault, D. Sacks, and M. Desjardins**. 2003. Phagosomes are competent organelles for antigen cross-presentation. Nature **425**(6956): 402-6.

138. **Houghton, R. L., M. J. Lodes, D. C. Dillon, L. D. Reynolds, C. H. Day, P. D. McNeill, R. C. Hendrickson, Y. A. Skeiky, D. P. Sampaio, R. Badaro, K. P. Lyashchenko, and S. G. Reed**. 2002. Use of multiepitope polyproteins in serodiagnosis of active tuberculosis. Clin Diagn Lab Immunol **9**(4): 883-91.

139. **Indrigo, J., R. L. Hunter, Jr., and J. K. Actor**. 2003. Cord factor trehalose 6,6'-dimycolate (TDM) mediates trafficking events during mycobacterial infection of murine macrophages. Microbiology **149**(Pt 8): 2049-59.

140. **Ioanoviciu, A., E. T. Yukl, P. Moenne-Loccoz, and P. R. de Montellano**. 2007. DevS, a heme-containing two-component oxygen sensor of *Mycobacterium tuberculosis*. Biochemistry **46**(14): 4250-60.

141. **Jackowski, S. and C. O. Rock**. 1986. Transfer of fatty acids from the 1-position of phosphatidylethanolamine to the major outer membrane lipoprotein of *Escherichia coli*. J Biol Chem **261**(24): 11328-33.

142. **Jeffery, C. J.** 1999. Moonlighting proteins. Trends Biochem Sci **24**(1): 8-11.

143. **Jo, E. K., C. S. Yang, C. H. Choi, and C. V. Harding**. 2007. Intracellular signalling cascades regulating innate immune responses to Mycobacteria: branching out from Toll-like receptors. Cell Microbiol **9**(5): 1087-98.

144. **Joe, M., D. Sun, H. Taha, G. C. Completo, J. E. Croudace, D. A. Lammas, G. S. Besra, and T. L. Lowary**. 2006. The 5-deoxy-5-methylthio-xylofuranose residue in mycobacterial lipoarabinomannan. absolute stereochemistry, linkage

position, conformation, and immunomodulatory activity. J Am Chem Soc **128**(15): 5059-72.

145. **Julian, E., L. Matas, A. Perez, J. Alcaide, M. A. Laneelle, and M. Luquin**. 2002. Serodiagnosis of tuberculosis: comparison of immunoglobulin A (IgA) response to sulfolipid I with IgG and IgM responses to 2,3-diacyltrehalose, 2,3,6-triacyltrehalose, and cord factor antigens. J Clin Microbiol **40**(10): 3782-8.

146. **Kang, C. M., D. W. Abbott, S. T. Park, C. C. Dascher, L. C. Cantley, and R. N. Husson**. 2005. The *Mycobacterium tuberculosis* serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape. Genes Dev **19**(14): 1692-704.

147. **Kapur, V., T. S. Whittam, and J. M. Musser**. 1994. Is *Mycobacterium tuberculosis* 15,000 years old? J Infect Dis **170**(5): 1348-9.

148. **Khoo, K. H., A. Dell, H. R. Morris, P. J. Brennan, and D. Chatterjee**. 1995. Structural definition of acylated phosphatidylinositol mannosides from *Mycobacterium tuberculosis*: definition of a common anchor for lipomannan and lipoarabinomannan. Glycobiology **5**(1): 117-27.

149. **Kotani, S., I. Yanagida, K. Kato, and T. Matsuda**. 1970. Studies on peptides, glycopeptides and antigenic polysaccharide-glycopeptide complexes isolated from an L-11 enzyme lysate of the cell walls of *Mycobacterium tuberculosis* strain H37Rv. Biken J **13**(4): 249-75.

150. **Koul, A., A. Choidas, M. Treder, A. K. Tyagi, K. Drlica, Y. Singh, and A. Ullrich**. 2000. Cloning and characterization of secretory tyrosine phosphatases of *Mycobacterium tuberculosis*. J Bacteriol **182**(19): 5425-32.

151. **Koul, A., T. Herget, B. Klebl, and A. Ullrich**. 2004. Interplay between mycobacteria and host signalling pathways. Nat Rev Microbiol **2**(3): 189-202.

152. **Laal, S.**, Immunodiagnosis, in Tuberculosis, S. Garay and W. Rom, Editors. 2004, Lippincott, Williams & Wilkins: Philadelphia. p. 185-91.

153. **Laal, S., K. M. Samanich, M. G. Sonnenberg, J. T. Belisle, J. O'Leary, M. S. Simberkoff, and S. Zolla-Pazner**. 1997. Surrogate marker of preclinical tuberculosis in human immunodeficiency virus infection: antibodies to an 88-kDa secreted antigen of *Mycobacterium tuberculosis*. J Infect Dis **176**(1): 133-43.

154. **Laal, S., K. M. Samanich, M. G. Sonnenberg, S. Zolla-Pazner, J. M. Phadtare, and J. T. Belisle**. 1997. Human humoral responses to antigens of *Mycobacterium tuberculosis*: immunodominance of high-molecular-mass antigens. Clin Diagn Lab Immunol **4**(1): 49-56.

155. **Laal, S. and Y. A. Skeiky**, Immune-Based Methods, in Tuberculosis and the Tubercle Bacillus, S.T. Cole, Editor. 2005, ASM Press: Washington, D.C. p. 71-83.

156. **Lalvani, A.** 2007. Diagnosing tuberculosis infection in the 21st century: new tools to tackle an old enemy. Chest **131**(6): 1898-906.

157. **Lara, M., L. Servin-Gonzalez, M. Singh, C. Moreno, I. Cohen, M. Nimtz, and C. Espitia**. 2004. Expression, secretion, and glycosylation of the 45- and 47-kDa glycoprotein of *Mycobacterium tuberculosis* in *Streptomyces lividans*. Appl Environ Microbiol **70**(2): 679-85.

158. **Lau, Y. L., G. C. Chan, S. Y. Ha, Y. F. Hui, and K. Y. Yuen**. 1998. The role of phagocytic respiratory burst in host defense against *Mycobacterium tuberculosis*. Clin Infect Dis **26**(1): 226-7.

159. **Lazarevic, V., D. Nolt, and J. L. Flynn**. 2005. Long-term control of *Mycobacterium tuberculosis* infection is mediated by dynamic immune responses. J Immunol **175**(2): 1107-17.

160. **Lee, S., B. Y. Jeon, S. Bardarov, M. Chen, S. L. Morris, and W. R. Jacobs, Jr.** 2006. Protection elicited by two glutamine auxotrophs of *Mycobacterium tuberculosis* and in vivo growth phenotypes of the four unique glutamine synthetase mutants in a murine model. Infect Immun **74**(11): 6491-5.

161. **Lefevre, P., O. Denis, L. De Wit, A. Tanghe, P. Vandenbussche, J. Content, and K. Huygen**. 2000. Cloning of the gene encoding a 22-kilodalton cell surface antigen of *Mycobacterium bovis* BCG and analysis of its potential for DNA vaccination against tuberculosis. Infect Immun **68**(3): 1040-7.

162. **Lefford, M. J.** 1975. Transfer of adoptive immunity to tuberculosis in mice. Infect Immun **11**(6): 1174-81.

163. **Lein, A. D., C. F. von Reyn, P. Ravn, C. R. Horsburgh, Jr., L. N. Alexander, and P. Andersen**. 1999. Cellular immune responses to ESAT-6 discriminate between patients with pulmonary disease due to *Mycobacterium avium* complex and those with pulmonary disease due to *Mycobacterium tuberculosis*. Clin Diagn Lab Immunol **6**(4): 606-9.

164. **Lemassu, A. and M. Daffe**. 1994. Structural features of the exocellular polysaccharides of *Mycobacterium tuberculosis*. Biochem J **297 (Pt 2)**: 351-7.

165. **Li, Y., E. Miltner, M. Wu, M. Petrofsky, and L. E. Bermudez**. 2005. A *Mycobacterium avium* PPE gene is associated with the ability of the bacterium to grow in macrophages and virulence in mice. Cell Microbiol **7**(4): 539-48.

166. **Li, Z., C. Kelley, F. Collins, D. Rouse, and S. Morris**. 1998. Expression of katG in *Mycobacterium tuberculosis* is associated with its growth and persistence in mice and guinea pigs. J Infect Dis **177**(4): 1030-5.

167. **Locht, C., J. M. Hougardy, C. Rouanet, S. Place, and F. Mascart**. 2006. Heparin-binding hemagglutinin, from an extrapulmonary dissemination factor to a powerful diagnostic and protective antigen against tuberculosis. Tuberculosis (Edinb) **86**(3-4): 303-9.

168. **Lodes, M. J., D. C. Dillon, R. Mohamath, C. H. Day, D. R. Benson, L. D. Reynolds, P. McNeill, D. P. Sampaio, Y. A. Skeiky, R. Badaro, D. H. Persing, S. G. Reed, and R. L. Houghton**. 2001. Serological expression cloning and immunological evaluation of MTB48, a novel *Mycobacterium tuberculosis* antigen. J Clin Microbiol **39**(7): 2485-93.

169. **Lowe, J. B. and J. D. Marth**. 2003. A genetic approach to Mammalian glycan function. Annu Rev Biochem **72**: 643-91.

170. **Ludwiczak, P., M. Gilleron, Y. Bordat, C. Martin, B. Gicquel, and G. Puzo**. 2002. *Mycobacterium tuberculosis* phoP mutant: lipoarabinomannan molecular structure. Microbiology **148**(Pt 10): 3029-37.

171. **Lyashchenko, K. P., M. Singh, R. Colangeli, and M. L. Gennaro**. 2000. A multi-antigen print immunoassay for the development of serological diagnosis of infectious diseases. J Immunol Methods **242**(1-2): 91-100.

172. **MacMicking, J., Q. W. Xie, and C. Nathan**. 1997. Nitric oxide and macrophage function. Annu Rev Immunol **15**: 323-50.

173. **MacMicking, J. D., R. J. North, R. LaCourse, J. S. Mudgett, S. K. Shah, and C. F. Nathan**. 1997. Identification of nitric oxide synthase as a protective locus against tuberculosis. Proc Natl Acad Sci U S A **94**(10): 5243-8.

174. **Maeda, N., J. Nigou, J. L. Herrmann, M. Jackson, A. Amara, P. H. Lagrange, G. Puzo, B. Gicquel, and O. Neyrolles**. 2003. The cell surface receptor DC-SIGN discriminates between Mycobacterium species through selective recognition of the mannose caps on lipoarabinomannan. J Biol Chem **278**(8): 5513-6.

175. **Mahapatra, S., H. Scherman, P. J. Brennan, and D. C. Crick**. 2005. N Glycolylation of the nucleotide precursors of peptidoglycan biosynthesis of Mycobacterium spp. is altered by drug treatment. J Bacteriol **187**(7): 2341-7.

176. **Manca, C., S. Paul, C. E. Barry, 3rd, V. H. Freedman, and G. Kaplan**. 1999. *Mycobacterium tuberculosis* catalase and peroxidase activities and resistance to oxidative killing in human monocytes in vitro. Infect Immun **67**(1): 74-9.

177. **Martin-Casabona, N., T. Gonzalez Fuente, F. Papa, J. Rossello Urgell, R. Vidal Pla, G. Codina Grau, and I. Ruiz Camps**. 1992. Time course of anti-SL-IV immunoglobulin G antibodies in patients with tuberculosis and tuberculosis-associated AIDS. J Clin Microbiol **30**(5): 1089-93.

178. **Martin-Orozco, N., N. Touret, M. L. Zaharik, E. Park, R. Kopelman, S. Miller, B. B. Finlay, P. Gros, and S. Grinstein**. 2006. Visualization of vacuolar acidification-induced transcription of genes of pathogens inside macrophages. Mol Biol Cell **17**(1): 498-510.

179. **Matsuba, T., Y. Suzuki, and Y. Tanaka**. 2007. Association of the Rv0679c protein with lipids and carbohydrates in *Mycobacterium tuberculosis/Mycobacterium bovis* BCG. Arch Microbiol **187**(4): 297-311.

180. **Mazzaccaro, R. J., M. Gedde, E. R. Jensen, H. M. van Santen, H. L. Ploegh, K. L. Rock, and B. R. Bloom**. 1996. Major histocompatibility class I presentation of soluble antigen facilitated by *Mycobacterium tuberculosis* infection. Proc Natl Acad Sci U S A **93**(21): 11786-91.

181. **McKinney, J. D., K. Honer zu Bentrup, E. J. Munoz-Elias, A. Miczak, B. Chen, W. T. Chan, D. Swenson, J. C. Sacchettini, W. R. Jacobs, Jr., and D. G. Russell**. 2000. Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature **406**(6797): 735-8.

182. **McNeil, M., M. Daffe, and P. J. Brennan**. 1990. Evidence for the nature of the link between the arabinogalactan and peptidoglycan of mycobacterial cell walls. J Biol Chem **265**(30): 18200-6.

183. **Mehta, R., J. T. Pearson, S. Mahajan, A. Nath, M. J. Hickey, D. R. Sherman, and W. M. Atkins**. 2004. Adenylylation and catalytic properties of *Mycobacterium tuberculosis* glutamine synthetase expressed in *Escherichia coli* versus mycobacteria. J Biol Chem **279**(21): 22477-82.

184. **Menozzi, F. D., R. Bischoff, E. Fort, M. J. Brennan, and C. Locht**. 1998. Molecular characterization of the mycobacterial heparin-binding hemagglutinin, a mycobacterial adhesin. Proc Natl Acad Sci U S A **95**(21): 12625-30.

185. **Menozzi, F. D., J. H. Rouse, M. Alavi, M. Laude-Sharp, J. Muller, R. Bischoff, M. J. Brennan, and C. Locht**. 1996. Identification of a heparin-binding hemagglutinin present in mycobacteria. J Exp Med **184**(3): 993-1001.

186. **Mescher, M. F. and J. L. Strominger**. 1976. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of Halobacterium salinarium. J Biol Chem **251**(7): 2005-14.

187. **Messner, P.** 2004. Prokaryotic glycoproteins: unexplored but important. J Bacteriol **186**(9): 2517-9.

188. **Michell, S. L., A. O. Whelan, P. R. Wheeler, M. Panico, R. L. Easton, A. T. Etienne, S. M. Haslam, A. Dell, H. R. Morris, A. J. Reason, J. L. Herrmann, D. B. Young, and R. G. Hewinson**. 2003. The MPB83 antigen from *Mycobacterium bovis* contains O-linked mannose and (1-->3)-mannobiose moieties. J Biol Chem **278**(18): 16423-32.

189. **Molle, V., A. K. Brown, G. S. Besra, A. J. Cozzone, and L. Kremer**. 2006. The condensing activities of the *Mycobacterium tuberculosis* type II fatty acid synthase are differentially regulated by phosphorylation. J Biol Chem **281**(40): 30094-103.

190. **Molle, V., L. Kremer, C. Girard-Blanc, G. S. Besra, A. J. Cozzone, and J. F. Prost**. 2003. An FHA phosphoprotein recognition domain mediates protein EmbR phosphorylation by PknH, a Ser/Thr protein kinase from *Mycobacterium tuberculosis*. Biochemistry **42**(51): 15300-9.

191. **Molle, V., D. Soulat, J. M. Jault, C. Grangeasse, A. J. Cozzone, and J. F. Prost**. 2004. Two FHA domains on an ABC transporter, Rv1747, mediate its phosphorylation by PknF, a Ser/Thr protein kinase from *Mycobacterium tuberculosis*. FEMS Microbiol Lett **234**(2): 215-23.

192. **Moody, D. B., M. R. Guy, E. Grant, T. Y. Cheng, M. B. Brenner, G. S. Besra, and S. A. Porcelli**. 2000. CD1b-mediated T cell recognition of a glycolipid antigen generated from mycobacterial lipid and host carbohydrate during infection. J Exp Med **192**(7): 965-76.

193. **Moody, D. B., B. B. Reinhold, M. R. Guy, E. M. Beckman, D. E. Frederique, S. T. Furlong, S. Ye, V. N. Reinhold, P. A. Sieling, R. L. Modlin, G. S. Besra, and S. A. Porcelli**. 1997. Structural requirements for glycolipid antigen recognition by CD1b-restricted T cells. Science **278**(5336): 283-6.

194. **Moody, D. B., T. Ulrichs, W. Muhlecker, D. C. Young, S. S. Gurcha, E. Grant, J. P. Rosat, M. B. Brenner, C. E. Costello, G. S. Besra, and S. A. Porcelli**. 2000. CD1c-mediated T-cell recognition of isoprenoid glycolipids in *Mycobacterium tuberculosis* infection. Nature **404**(6780): 884-8.

195. **Moody, D. B., D. C. Young, T. Y. Cheng, J. P. Rosat, C. Roura-Mir, P. B. O'Connor, D. M. Zajonc, A. Walz, M. J. Miller, S. B. Levery, I. A. Wilson, C. E. Costello, and M. B. Brenner**. 2004. T cell activation by lipopeptide antigens. Science **303**(5657): 527-31.

196. **Morth, J. P., S. Gosmann, E. Nowak, and P. A. Tucker**. 2005. A novel two-component system found in *Mycobacterium tuberculosis*. FEBS Lett **579**(19): 4145-8.

197. **Mueller-Ortiz, S. L., A. R. Wanger, and S. J. Norris**. 2001. Mycobacterial protein HbhA binds human complement component C3. Infect Immun **69**(12): 7501-11.

198. **Nagai, S., H. G. Wiker, M. Harboe, and M. Kinomoto**. 1991. Isolation and partial characterization of major protein antigens in the culture fluid of *Mycobacterium tuberculosis*. Infect Immun **59**(1): 372-82.

199. **Narita, S., S. Matsuyama, and H. Tokuda**. 2004. Lipoprotein trafficking in *Escherichia coli*. Arch Microbiol **182**(1): 1-6.

200. **Nicholson, S., G. Bonecini-Almeida Mda, J. R. Lapa e Silva, C. Nathan, Q. W. Xie, R. Mumford, J. R. Weidner, J. Calaycay, J. Geng, N. Boechat, C. Linhares, W. Rom, and J. L. Ho**. 1996. Inducible nitric oxide synthase in pulmonary alveolar macrophages from patients with tuberculosis. J Exp Med **183**(5): 2293-302.

201. **Nielsen, J. B. and J. O. Lampen**. 1982. Glyceride-cysteine lipoproteins and secretion by Gram-positive bacteria. J Bacteriol **152**(1): 315-22.

202. **Nigou, J., M. Gilleron, and G. Puzo**. 2003. Lipoarabinomannans: from structure to biosynthesis. Biochimie **85**(1-2): 153-66.

203. **Okkels, L. M., E. C. Muller, M. Schmid, I. Rosenkrands, S. H. Kaufmann, P. Andersen, and P. R. Jungblut**. 2004. CFP10 discriminates between nonacetylated and acetylated ESAT-6 of *Mycobacterium tuberculosis* by differential interaction. Proteomics **4**(10): 2954-60.

204. **Orme, I. M. and F. M. Collins**. 1983. Protection against *Mycobacterium tuberculosis* infection by adoptive immunotherapy. Requirement for T cell-deficient recipients. J Exp Med **158**(1): 74-83.

205. **Ortalo-Magne, A., M. A. Dupont, A. Lemassu, A. B. Andersen, P. Gounon, and M. Daffe**. 1995. Molecular composition of the outermost capsular material of the tubercle bacillus. Microbiology **141 (Pt 7)**: 1609-20.

206. **Pallen, M. J., A. C. Lam, N. J. Loman, and A. McBride**. 2001. An abundance of bacterial ADP-ribosyltransferases--implications for the origin of exotoxins and their human homologues. Trends Microbiol **9**(7): 302-7; discussion 8.

207. **Pang, X. and S. T. Howard**. 2007. Regulation of the {alpha}-Crystallin Gene acr2 by the MprAB Two-Component System of *Mycobacterium tuberculosis*. J Bacteriol **189**(17): 6213-21.

208. **Parish, T. and N. G. Stoker**. 2000. glnE is an essential gene in *Mycobacterium tuberculosis*. J Bacteriol **182**(20): 5715-20.

209. **Park, H. D., K. M. Guinn, M. I. Harrell, R. Liao, M. I. Voskuil, M. Tompa, G. K. Schoolnik, and D. R. Sherman**. 2003. Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. Mol Microbiol **48**(3): 833-43.

210. **Pashley, C. A., A. C. Brown, D. Robertson, and T. Parish**. 2006. Identification of the *Mycobacterium tuberculosis* GlnE promoter and its response to nitrogen availability. Microbiology **152**(Pt 9): 2727-34.

211. **Pasula, R., J. F. Downing, J. R. Wright, D. L. Kachel, T. E. Davis, Jr., and W. J. Martin, 2nd**. 1997. Surfactant protein A (SP-A) mediates attachment of *Mycobacterium tuberculosis* to murine alveolar macrophages. Am J Respir Cell Mol Biol **17**(2): 209-17.

212. **Pecora, N. D., A. J. Gehring, D. H. Canaday, W. H. Boom, and C. V. Harding**. 2006. *Mycobacterium tuberculosis* LprA is a lipoprotein agonist of TLR2 that regulates innate immunity and APC function. J Immunol **177**(1): 422-9.

213. **Perkins, M. D., G. Roscigno, and A. Zumla**. 2006. Progress towards improved tuberculosis diagnostics for developing countries. Lancet **367**(9514): 942-3.

214. **Pethe, K., S. Alonso, F. Biet, G. Delogu, M. J. Brennan, C. Locht, and F. D. Menozzi**. 2001. The heparin-binding haemagglutinin of M. tuberculosis is required for extrapulmonary dissemination. Nature **412**(6843): 190-4.

215. **Pethe, K., M. Aumercier, E. Fort, C. Gatot, C. Locht, and F. D. Menozzi**. 2000. Characterization of the heparin-binding site of the mycobacterial heparin-binding hemagglutinin adhesin. J Biol Chem **275**(19): 14273-80.

216. **Pethe, K., P. Bifani, H. Drobecq, C. Sergheraert, A. S. Debrie, C. Locht, and F. D. Menozzi**. 2002. Mycobacterial heparin-binding hemagglutinin and laminin-binding protein share antigenic methyllysines that confer resistance to proteolysis. Proc Natl Acad Sci U S A **99**(16): 10759-64.

217. **Pethe, K., V. Puech, M. Daffe, C. Josenhans, H. Drobecq, C. Locht, and F. D. Menozzi**. 2001. *Mycobacterium smegmatis* laminin-binding glycoprotein shares epitopes with *Mycobacterium tuberculosis* heparin-binding haemagglutinin. Mol Microbiol **39**(1): 89-99.

218. **Philips, J. A., E. J. Rubin, and N. Perrimon**. 2005. Drosophila RNAi screen reveals CD36 family member required for mycobacterial infection. Science **309**(5738): 1251-3.

219. **Piddington, D. L., F. C. Fang, T. Laessig, A. M. Cooper, I. M. Orme, and N. A. Buchmeier**. 2001. Cu,Zn superoxide dismutase of *Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst. Infect Immun **69**(8): 4980-7.

220. **Pollock, J. M. and P. Andersen**. 1997. Predominant recognition of the ESAT-6 protein in the first phase of interferon with *Mycobacterium bovis* in cattle. Infect Immun **65**(7): 2587-92.

221. **Portevin, D., C. De Sousa-D'Auria, C. Houssin, C. Grimaldi, M. Chami, M. Daffe, and C. Guilhot**. 2004. A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. Proc Natl Acad Sci U S A **101**(1): 314-9.

222. **Pullen, K. E., H. L. Ng, P. Y. Sung, M. C. Good, S. M. Smith, and T. Alber**. 2004. An alternate conformation and a third metal in PstP/Ppp, the M. tuberculosis PP2C-Family Ser/Thr protein phosphatase. Structure **12**(11): 1947-54.

223. **Ragas, A., L. Roussel, G. Puzo, and M. Riviere**. 2007. The *Mycobacterium tuberculosis* cell-surface glycoprotein apa as a potential adhesin to colonize target cells via the innate immune system pulmonary C-type lectin surfactant protein A. J Biol Chem **282**(8): 5133-42.

224. **Ramakrishnan, L., N. A. Federspiel, and S. Falkow**. 2000. Granuloma-specific expression of Mycobacterium virulence proteins from the glycine-rich PE-PGRS family. Science **288**(5470): 1436-9.

225. **Raynaud, C., G. Etienne, P. Peyron, M. A. Laneelle, and M. Daffe**. 1998. Extracellular enzyme activities potentially involved in the pathogenicity of *Mycobacterium tuberculosis*. Microbiology **144 (Pt 2)**: 577-87.

226. **Read, R., C. A. Pashley, D. Smith, and T. Parish**. 2007. The role of GlnD in ammonia assimilation in *Mycobacterium tuberculosis*. Tuberculosis (Edinb) **87**(4): 384-90.

227. **Rhee, K. Y., H. Erdjument-Bromage, P. Tempst, and C. F. Nathan**. 2005. S-nitroso proteome of *Mycobacterium tuberculosis*: Enzymes of intermediary metabolism and antioxidant defense. Proc Natl Acad Sci U S A **102**(2): 467-72.

228. **Rich, E. A., M. Torres, E. Sada, C. K. Finegan, B. D. Hamilton, and Z. Toossi**. 1997. *Mycobacterium tuberculosis* (MTB)-stimulated production of nitric oxide by human alveolar macrophages and relationship of nitric oxide production to growth inhibition of MTB. Tuber Lung Dis **78**(5-6): 247-55.

229. **Riley, R. L., C. C. Mills, and W. Nyka**. 1959. Aerial dissemination of pulmonary tuberculosis. Am J Hyg **70**: 185-96.

230. **Rison, S. C., J. Mattow, P. R. Jungblut, and N. G. Stoker**. 2007. Experimental determination of translational starts using peptide mass mapping and tandem mass spectrometry within the proteome of *Mycobacterium tuberculosis*. Microbiology **153**(Pt 2): 521-8.

231. **Roberts, D. M., R. P. Liao, G. Wisedchaisri, W. G. Hol, and D. R. Sherman**. 2004. Two sensor kinases contribute to the hypoxic response of *Mycobacterium tuberculosis*. J Biol Chem **279**(22): 23082-7.

232. **Romain, F., C. Horn, P. Pescher, A. Namane, M. Riviere, G. Puzo, O. Barzu, and G. Marchal.** 1999. Deglycosylation of the 45/47-kilodalton antigen complex of *Mycobacterium tuberculosis* decreases its capacity to elicit in vivo or in vitro cellular immune responses. Infect Immun **67**(11): 5567-72.

233. **Ruan, J., G. St John, S. Ehrt, L. Riley, and C. Nathan.** 1999. noxR3, a novel gene from *Mycobacterium tuberculosis*, protects Salmonella typhimurium from nitrosative and oxidative stress. Infect Immun **67**(7): 3276-83.

234. **Russell, D. G.** 1995. Mycobacterium and Leishmania: stowaways in the endosomal network. Trends Cell Biol **5**(3): 125-8.

235. **Russell, D. G.** 2007. Who puts the tubercle in tuberculosis? Nat Rev Microbiol **5**(1): 39-47.

236. **Sada, E., L. E. Ferguson, and T. M. Daniel.** 1990. An ELISA for the serodiagnosis of tuberculosis using a 30,000-Da native antigen of *Mycobacterium tuberculosis*. J Infect Dis **162**(4): 928-31.

237. **Saini, D. K., V. Malhotra, D. Dey, N. Pant, T. K. Das, and J. S. Tyagi.** 2004. DevR-DevS is a bona fide two-component system of *Mycobacterium tuberculosis* that is hypoxia-responsive in the absence of the DNA-binding domain of DevR. Microbiology **150**(Pt 4): 865-75.

238. **Salo, W. L., A. C. Aufderheide, J. Buikstra, and T. A. Holcomb.** 1994. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. Proc Natl Acad Sci U S A **91**(6): 2091-4.

239. **Samanich, K., J. T. Belisle, and S. Laal.** 2001. Homogeneity of antibody responses in tuberculosis patients. Infect Immun **69**(7): 4600-9.

240. **Samanich, K. M., J. T. Belisle, M. G. Sonnenberg, M. A. Keen, S. Zolla-Pazner, and S. Laal.** 1998. Delineation of human antibody responses to culture filtrate antigens of *Mycobacterium tuberculosis*. J Infect Dis **178**(5): 1534-8.

241. **Samanich, K. M., M. A. Keen, V. D. Vissa, J. D. Harder, J. S. Spencer, J. T. Belisle, S. Zolla-Pazner, and S. Laal.** 2000. Serodiagnostic potential of culture filtrate antigens of *Mycobacterium tuberculosis*. Clin Diagn Lab Immunol **7**(4): 662-8.

242. **Samols, D., C. G. Thornton, V. L. Murtif, G. K. Kumar, F. C. Haase, and H. G. Wood.** 1988. Evolutionary conservation among biotin enzymes. J Biol Chem **263**(14): 6461-4.

243. **Sander, P., M. Rezwan, B. Walker, S. K. Rampini, R. M. Kroppenstedt, S. Ehlers, C. Keller, J. R. Keeble, M. Hagemeier, M. J. Colston, B. Springer, and E. C. Bottger.** 2004. Lipoprotein processing is required for virulence of *Mycobacterium tuberculosis*. Mol Microbiol **52**(6): 1543-52.

244. **Sankaran, K. and H. C. Wu**. 1994. Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. J Biol Chem **269**(31): 19701-6.

245. **Sassetti, C. M., D. H. Boyd, and E. J. Rubin**. 2001. Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci U S A **98**(22): 12712-7.

246. **Scanga, C. A., V. P. Mohan, K. Tanaka, D. Alland, J. L. Flynn, and J. Chan**. 2001. The inducible nitric oxide synthase locus confers protection against aerogenic challenge of both clinical and laboratory strains of *Mycobacterium tuberculosis* in mice. Infect Immun **69**(12): 7711-7.

247. **Schleifer, K. H. and O. Kandler**. 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol Rev **36**(4): 407-77.

248. **Schlesinger, L. S.** 1993. Macrophage phagocytosis of virulent but not attenuated strains of *Mycobacterium tuberculosis* is mediated by mannose receptors in addition to complement receptors. J Immunol **150**(7): 2920-30.

249. **Schlesinger, L. S.**, Phagocytosis and Toll-like receptors in tuberculosis, in Tuberculosis, S. Garay and W.N. Rom, Editors. 2004, Lippincott Williams & Wilkins: Philadelphia. p. 203-14.

250. **Schlesinger, L. S., C. G. Bellinger-Kawahara, N. R. Payne, and M. A. Horwitz**. 1990. Phagocytosis of *Mycobacterium tuberculosis* is mediated by human monocyte complement receptors and complement component C3. J Immunol **144**(7): 2771-80.

251. **Schlesinger, L. S., S. R. Hull, and T. M. Kaufman**. 1994. Binding of the terminal mannosyl units of lipoarabinomannan from a virulent strain of *Mycobacterium tuberculosis* to human macrophages. J Immunol **152**(8): 4070-9.

252. **Schmidt, M. A., L. W. Riley, and I. Benz**. 2003. Sweet new world: glycoproteins in bacterial pathogens. Trends Microbiol **11**(12): 554-61.

253. **Serres, M. H. and J. C. Ensign**. 1996. Endogenous ADP-ribosylation of proteins in *Mycobacterium smegmatis*. J Bacteriol **178**(20): 6074-7.

254. **Shapiro, B. M., H. S. Kingdon, and E. R. Stadtman**. 1967. Regulation of glutamine synthetase. VII. Adenylyl glutamine synthetase: a new form of the enzyme with altered regulatory and kinetic properties. Proc Natl Acad Sci U S A **58**(2): 642-9.

255. **Shapiro, B. M. and E. R. Stadtman**. 1968. 5'-adenylyl-O-tyrosine. The novel phosphodiester residue of adenylylated glutamine synthetase from *Escherichia coli*. J Biol Chem **243**(13): 3769-71.

256. **Sharma, K., M. Gupta, A. Krupa, N. Srinivasan, and Y. Singh**. 2006. EmbR, a regulatory protein with ATPase activity, is a substrate of multiple serine/threonine kinases and phosphatase in *Mycobacterium tuberculosis*. Febs J **273**(12): 2711-21.

257. **Sherman, D. R., M. Voskuil, D. Schnappinger, R. Liao, M. I. Harrell, and G. K. Schoolnik**. 2001. Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha -crystallin. Proc Natl Acad Sci U S A **98**(13): 7534-9.

258. **Shi, L., Y. J. Jung, S. Tyagi, M. L. Gennaro, and R. J. North**. 2003. Expression of Th1-mediated immunity in mouse lungs induces a *Mycobacterium tuberculosis* transcription pattern characteristic of nonreplicating persistence. Proc Natl Acad Sci U S A **100**(1): 241-6.

259. **Shin, A. R., K. S. Lee, J. S. Lee, S. Y. Kim, C. H. Song, S. B. Jung, C. S. Yang, E. K. Jo, J. K. Park, T. H. Paik, and H. J. Kim**. 2006. *Mycobacterium tuberculosis* HBHA protein reacts strongly with the serum immunoglobulin M of tuberculosis patients. Clin Vaccine Immunol **13**(8): 869-75.

260. **Shingadia, D. and V. Novelli**. 2003. Diagnosis and treatment of tuberculosis in children. Lancet Infect Dis **3**(10): 624-32.

261. **Sickmann, A. and H. E. Meyer**. 2001. Phosphoamino acid analysis. Proteomics **1**(2): 200-6.

262. **Sidobre, S., J. Nigou, G. Puzo, and M. Riviere**. 2000. Lipoglycans are putative ligands for the human pulmonary surfactant protein A attachment to mycobacteria. Critical role of the lipids for lectin-carbohydrate recognition. J Biol Chem **275**(4): 2415-22.

263. **Sidobre, S., G. Puzo, and M. Riviere**. 2002. Lipid-restricted recognition of mycobacterial lipoglycans by human pulmonary surfactant protein A: a surface-plasmon-resonance study. Biochem J **365**(Pt 1): 89-97.

264. **Sieling, P. A., D. Chatterjee, S. A. Porcelli, T. I. Prigozy, R. J. Mazzaccaro, T. Soriano, B. R. Bloom, M. B. Brenner, M. Kronenberg, P. J. Brennan, and et al.** 1995. CD1-restricted T cell recognition of microbial lipoglycan antigens. Science **269**(5221): 227-30.

265. **Singh, K. K., Y. Dong, J. T. Belisle, J. Harder, V. K. Arora, and S. Laal**. 2005. Antigens of *Mycobacterium tuberculosis* recognized by antibodies during incipient, subclinical tuberculosis. Clin Diagn Lab Immunol **12**(2): 354-8.

266. **Singh, K. K., Y. Dong, L. Hinds, M. A. Keen, J. T. Belisle, S. Zolla-Pazner, J. M. Achkar, A. J. Nadas, V. K. Arora, and S. Laal**. 2003. Combined use of serum and urinary antibody for diagnosis of tuberculosis. J Infect Dis **188**(3): 371-7.

267. **Singh, K. K., Y. Dong, S. A. Patibandla, D. N. McMurray, V. K. Arora, and S. Laal**. 2005. Immunogenicity of the *Mycobacterium tuberculosis* PPE55 (Rv3347c) protein during incipient and clinical tuberculosis. Infect Immun **73**(8): 5004-14.

268. **Singh, K. K., X. Zhang, A. S. Patibandla, P. Chien, Jr., and S. Laal**. 2001. Antigens of *Mycobacterium tuberculosis* expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences. Infect Immun **69**(6): 4185-91.

269. **Skeiky, Y. A., P. J. Ovendale, S. Jen, M. R. Alderson, D. C. Dillon, S. Smith, C. B. Wilson, I. M. Orme, S. G. Reed, and A. Campos-Neto**. 2000. T cell expression cloning of a *Mycobacterium tuberculosis* gene encoding a protective antigen associated with the early control of infection. J Immunol **165**(12): 7140-9.

270. **Sorensen, A. L., S. Nagai, G. Houen, P. Andersen, and A. B. Andersen**. 1995. Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. Infect Immun **63**(5): 1710-7.

271. **Sousa, E. H., J. R. Tuckerman, G. Gonzalez, and M. A. Gilles-Gonzalez**. 2007. DosT and DevS are oxygen-switched kinases in *Mycobacterium tuberculosis*. Protein Sci **16**(8): 1708-19.

272. **Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser**. 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A **94**(18): 9869-74.

273. **St John, G., N. Brot, J. Ruan, H. Erdjument-Bromage, P. Tempst, H. Weissbach, and C. Nathan**. 2001. Peptide methionine sulfoxide reductase from *Escherichia coli* and *Mycobacterium tuberculosis* protects bacteria against oxidative damage from reactive nitrogen intermediates. Proc Natl Acad Sci U S A **98**(17): 9901-6.

274. **Steenken, W., W. H. Oatway, and S. A. Petroff**. 1934. Biological studies of the tubercle bacillus. III Dissociation and pathogenecity of the human tubercle bacillus (H37). J Exp Med **60**: 515-43.

275. **Steingart, K. R., M. Henry, S. Laal, P. C. Hopewell, A. Ramsay, D. Menzies, J. Cunningham, K. Weldingh, and M. Pai**. 2007. Commercial serological antibody detection tests for the diagnosis of pulmonary tuberculosis: a systematic review. PLoS Med **4**(6): e202.

276. **Steingart, K. R., V. Ng, M. Henry, P. C. Hopewell, A. Ramsay, J. Cunningham, R. Urbanczik, M. D. Perkins, M. A. Aziz, and M. Pai**. 2006. Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. Lancet Infect Dis **6**(10): 664-74.

277. **Stenger, S., D. A. Hanson, R. Teitelbaum, P. Dewan, K. R. Niazi, C. J. Froelich, T. Ganz, S. Thoma-Uszynski, A. Melian, C. Bogdan, S. A. Porcelli, B. R. Bloom, A. M. Krensky, and R. L. Modlin**. 1998. An antimicrobial activity of cytolytic T cells mediated by granulysin. Science **282**(5386): 121-5.

278. **Stenger, S., R. J. Mazzaccaro, K. Uyemura, S. Cho, P. F. Barnes, J. P. Rosat, A. Sette, M. B. Brenner, S. A. Porcelli, B. R. Bloom, and R. L. Modlin**. 1997. Differential effects of cytolytic T cell subsets on intracellular infection. Science **276**(5319): 1684-7.

279. **Stock, A. M., V. L. Robinson, and P. N. Goudreau**. 2000. Two-component signal transduction. Annu Rev Biochem **69**: 183-215.

280. **Stokes, R. W., I. D. Haidl, W. A. Jefferies, and D. P. Speert**. 1993. Mycobacteria-macrophage interactions. Macrophage phenotype determines the nonopsonic binding of *Mycobacterium tuberculosis* to murine macrophages. J Immunol **151**(12): 7067-76.

281. **Sturgill-Koszycki, S., P. H. Schlesinger, P. Chakraborty, P. L. Haddix, H. L. Collins, A. K. Fok, R. D. Allen, S. L. Gluck, J. Heuser, and D. G. Russell**. 1994. Lack of acidification in Mycobacterium phagosomes produced by exclusion of the vesicular proton-ATPase. Science **263**(5147): 678-81.

282. **Sulzenbacher, G., S. Canaan, Y. Bordat, O. Neyrolles, G. Stadthagen, V. Roig-Zamboni, J. Rauzier, D. Maurin, F. Laval, M. Daffe, C. Cambillau, B. Gicquel, Y. Bourne, and M. Jackson**. 2006. LppX is a lipoprotein required for the translocation of phthiocerol dimycocerosates to the surface of *Mycobacterium tuberculosis*. Embo J **25**(7): 1436-44.

283. **Sutcliffe, I. C. and D. J. Harrington**. 2002. Pattern searches for the identification of putative lipoprotein genes in Gram-positive bacterial genomes. Microbiology **148**(Pt 7): 2065-77.

284. **Sutcliffe, I. C. and D. J. Harrington**. 2004. Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. FEMS Microbiol Rev **28**(5): 645-59.

285. **Sutcliffe, I. C. and R. R. Russell**. 1995. Lipoproteins of gram-positive bacteria. J Bacteriol **177**(5): 1123-8.

286. **Taylor, G. M., M. Goyal, A. J. Legge, R. J. Shaw, and D. Young**. 1999. Genotypic analysis of *Mycobacterium tuberculosis* from medieval human remains. Microbiology **145 (Pt 4)**: 899-904.

287. **Taylor, G. M., E. Murphy, R. Hopkins, P. Rutland, and Y. Chistov**. 2007. First report of *Mycobacterium bovis* DNA in human remains from the Iron Age. Microbiology **153**(Pt 4): 1243-9.

288. **Teitelbaum, R., M. Cammer, M. L. Maitland, N. E. Freitag, J. Condeelis, and B. R. Bloom**. 1999. Mycobacterial infection of macrophages results in membrane-permeable phagosomes. Proc Natl Acad Sci U S A **96**(26): 15190-5.

289. **Temmerman, S., K. Pethe, M. Parra, S. Alonso, C. Rouanet, T. Pickett, A. Drowart, A. S. Debrie, G. Delogu, F. D. Menozzi, C. Sergheraert, M. J. Brennan, F. Mascart, and C. Locht**. 2004. Methylation-dependent T cell immunity to *Mycobacterium tuberculosis* heparin-binding hemagglutinin. Nat Med **10**(9): 935-41.

290. **Tessema, T. A., G. Bjune, B. Hamasur, S. Svenson, H. Syre, and B. Bjorvatn**. 2002. Circulating antibodies to lipoarabinomannan in relation to sputum microscopy, clinical features and urinary anti-lipoarabinomannan detection in pulmonary tuberculosis. Scand J Infect Dis **34**(2): 97-103.

291. **Tjalsma, H. and J. M. van Dijl**. 2005. Proteomics-based consensus prediction of protein retention in a bacterial membrane. Proteomics **5**(17): 4472-82.

292. **Treumann, A., F. Xidong, L. McDonnell, P. J. Derrick, A. E. Ashcroft, D. Chatterjee, and S. W. Homans**. 2002. 5-Methylthiopentose: a new substituent on lipoarabinomannan in *Mycobacterium tuberculosis*. J Mol Biol **316**(1): 89-100.

293. **Tsai, M. C., S. Chakravarty, G. Zhu, J. Xu, K. Tanaka, C. Koch, J. Tufariello, J. Flynn, and J. Chan**. 2006. Characterization of the tuberculous granuloma in murine and human lungs: cellular composition and relative tissue oxygen tension. Cell Microbiol **8**(2): 218-32.

294. **Tullius, M. V., G. Harth, and M. A. Horwitz**. 2003. Glutamine synthetase GlnA1 is essential for growth of *Mycobacterium tuberculosis* in human THP-1 macrophages and guinea pigs. Infect Immun **71**(7): 3927-36.

295. **Turnbull, W. B., K. H. Shimizu, D. Chatterjee, S. W. Homans, and A. Treumann**. 2004. Identification of the 5-methylthiopentosyl substituent in *Mycobacterium tuberculosis* lipoarabinomannan. Angew Chem Int Ed Engl **43**(30): 3918-22.

296. **Turner, J., C. D. D'Souza, J. E. Pearl, P. Marietta, M. Noel, A. A. Frank, R. Appelberg, I. M. Orme, and A. M. Cooper**. 2001. CD8- and CD95/95L-dependent mechanisms of resistance in mice with chronic pulmonary tuberculosis. Am J Respir Cell Mol Biol **24**(2): 203-9.

297. **Upreti, R. K., M. Kumar, and V. Shankar**. 2003. Bacterial glycoproteins: functions, biosynthesis and applications. Proteomics **3**(4): 363-79.

298. **van der Wel, N., D. Hava, D. Houben, D. Fluitsma, M. van Zon, J. Pierson, M. Brenner, and P. J. Peters**. 2007. M. tuberculosis and M. leprae translocate from the phagolysosome to the cytosol in myeloid cells. Cell **129**(7): 1287-98.

299. **van Pinxteren, L. A., P. Ravn, E. M. Agger, J. Pollock, and P. Andersen**. 2000. Diagnosis of tuberculosis based on the two specific antigens ESAT-6 and CFP10. Clin Diagn Lab Immunol **7**(2): 155-60.

300. **VanderVen, B. C., J. D. Harder, D. C. Crick, and J. T. Belisle**. 2005. Export-mediated assembly of mycobacterial glycoproteins parallels eukaryotic pathways. Science **309**(5736): 941-3.

301. **Vergne, I., J. Chua, and V. Deretic**. 2003. Tuberculosis toxin blocking phagosome maturation inhibits a novel Ca2+/calmodulin-PI3K hVPS34 cascade. J Exp Med **198**(4): 653-9.

302. **Vergne, I., J. Chua, H. H. Lee, M. Lucas, J. Belisle, and V. Deretic**. 2005. Mechanism of phagolysosome biogenesis block by viable *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A **102**(11): 4033-8.

303. **Via, L. E., R. Curcic, M. H. Mudd, S. Dhandayuthapani, R. J. Ulmer, and V. Deretic**. 1996. Elements of signal transduction in *Mycobacterium tuberculosis*: in vitro phosphorylation and in vivo expression of the response regulator MtrA. J Bacteriol **178**(11): 3314-21.

304. **Via, L. E., D. Deretic, R. J. Ulmer, N. S. Hibler, L. A. Huber, and V. Deretic**. 1997. Arrest of mycobacterial phagosome maturation is caused by a block in vesicle fusion between stages controlled by rab5 and rab7. J Biol Chem **272**(20): 13326-31.

305. **von Heijne, G.** 1989. The structure of signal peptides from bacterial lipoproteins. Protein Eng **2**(7): 531-4.

306. **Voskuil, M. I., D. Schnappinger, K. C. Visconti, M. I. Harrell, G. M. Dolganov, D. R. Sherman, and G. K. Schoolnik**. 2003. Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. J Exp Med **198**(5): 705-13.

307. **Vosloo, W., P. Tippoo, J. E. Hughes, N. Harriman, M. Emms, D. W. Beatty, H. Zappe, and L. M. Steyn**. 1997. Characterisation of a lipoprotein in *Mycobacterium bovis* (BCG) with sequence similarity to the secreted protein MPB70. Gene **188**(1): 123-8.

308. **Walburger, A., A. Koul, G. Ferrari, L. Nguyen, C. Prescianotto-Baschong, K. Huygen, B. Klebl, C. Thompson, G. Bacher, and J. Pieters**. 2004. Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. Science **304**(5678): 1800-4.

309. **Walters, S. B., E. Dubnau, I. Kolesnikova, F. Laval, M. Daffe, and I. Smith**. 2006. The *Mycobacterium tuberculosis* PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. Mol Microbiol **60**(2): 312-30.

310. **Wang, C. H., C. Y. Liu, H. C. Lin, C. T. Yu, K. F. Chung, and H. P. Kuo.** 1998. Increased exhaled nitric oxide in active pulmonary tuberculosis due to inducible NO synthase upregulation in alveolar macrophages. Eur Respir J **11**(4): 809-15.

311. **Weldingh, K., I. Rosenkrands, S. Jacobsen, P. B. Rasmussen, M. J. Elhay, and P. Andersen.** 1998. Two-dimensional electrophoresis for analysis of *Mycobacterium tuberculosis* culture filtrate and purification and characterization of six novel proteins. Infect Immun **66**(8): 3492-500.

312. **Wilcke, J. T., B. N. Jensen, P. Ravn, A. B. Andersen, and K. Haslov.** 1996. Clinical evaluation of MPT-64 and MPT-59, two proteins secreted from *Mycobacterium tuberculosis*, for skin test reagents. Tuber Lung Dis **77**(3): 250-6.

313. **Wu, C. H., J. J. Tsai-Wu, Y. T. Huang, C. Y. Lin, G. G. Lioua, and F. J. Lee.** 1998. Identification and subcellular localization of a novel Cu,Zn superoxide dismutase of *Mycobacterium tuberculosis*. FEBS Lett **439**(1-2): 192-6.

314. **Yamagata, H., K. Daishima, and S. Mizushima.** 1983. Cloning and expression of a gene coding for the prolipoprotein signal peptidase of *Escherichia coli*. FEBS Lett **158**(2): 301-4.

315. **Young, D. B. and T. R. Garbe.** 1991. Lipoprotein antigens of *Mycobacterium tuberculosis*. Res Microbiol **142**(1): 55-65.

316. **Young, D. B. and J. R. Lamb.** 1986. T lymphocytes respond to solid-phase antigen: a novel approach to the molecular analysis of cellular immunity. Immunology **59**(2): 167-71.

317. **Yuan, Y., D. D. Crane, and C. E. Barry, 3rd.** 1996. Stationary phase-associated protein expression in *Mycobacterium tuberculosis*: function of the mycobacterial alpha-crystallin homolog. J Bacteriol **178**(15): 4484-92.

318. **Yuan, Y., D. D. Crane, R. M. Simpson, Y. Q. Zhu, M. J. Hickey, D. R. Sherman, and C. E. Barry, 3rd.** 1998. The 16-kDa alpha-crystallin (Acr) protein of *Mycobacterium tuberculosis* is required for growth in macrophages. Proc Natl Acad Sci U S A **95**(16): 9578-83.

319. **Zahrt, T. C., C. Wozniak, D. Jones, and A. Trevett.** 2003. Functional analysis of the *Mycobacterium tuberculosis* MprAB two-component signal transduction system. Infect Immun **71**(12): 6962-70.

320. **Zanetti, S., A. Bua, G. Delogu, C. Pusceddu, M. Mura, F. Saba, P. Pirina, C. Garzelli, C. Vertuccio, L. A. Sechi, and G. Fadda.** 2005. Patients with pulmonary tuberculosis develop a strong humoral response against methylated heparin-binding hemagglutinin. Clin Diagn Lab Immunol **12**(9): 1135-8.

321. **Zeng, Z., A. R. Castano, B. W. Segelke, E. A. Stura, P. A. Peterson, and I. A. Wilson**. 1997. Crystal structure of mouse CD1: An MHC-like fold with a large hydrophobic binding groove. Science **277**(5324): 339-45.

322. **Zhang, Y., R. Lathigra, T. Garbe, D. Catty, and D. Young**. 1991. Genetic analysis of superoxide dismutase, the 23 kilodalton antigen of *Mycobacterium tuberculosis*. Mol Microbiol **5**(2): 381-91.

323. **Zheng, X., K. G. Papavinasasundaram, and Y. Av-Gay**. 2007. Novel substrates of *Mycobacterium tuberculosis* PknH Ser/Thr kinase. Biochem Biophys Res Commun **355**(1): 162-8.

324. **Zimmerli, S., S. Edwards, and J. D. Ernst**. 1996. Selective receptor blockade during phagocytosis does not alter the survival and growth of *Mycobacterium tuberculosis* in human macrophages. Am J Respir Cell Mol Biol **15**(6): 760-70.

325. **Zink, A. R., C. Sola, U. Reischl, W. Grabner, N. Rastogi, H. Wolf, and A. G. Nerlich**. 2003. Characterization of *Mycobacterium tuberculosis* complex DNAs from Egyptian mummies by spoligotyping. J Clin Microbiol **41**(1): 359-67.

Chapter II

Literature Review Part II: Proteomic Approaches to Biomarker Discovery

2.1 Introduction

The term "biomarker" can have many different meanings that are often dependent on the context used. In a broad sense, a biomarker is a molecule that reflects a biologic state. To provide focus this chapter will address only proteins or peptides that are associated with the absence or presence of a specific disease. Recent years have witnessed an intense interest and significant advances in the discovery and evaluation of such molecules through proteomic methodologies, and the majority of these strategies were being developed and published during the course of this dissertation work. This chapter describes the biological sources, experimental strategies, and challenges for biomarker discovery, and further defines the current state of TB biomarkers.

2.2 The biofluid proteome

Biofluids can be analyzed for disease diagnosis and prognosis, as they are capable of reflecting the human physiological or pathological state (54). The study of biofluids offers several advantages over direct analysis of the localized diseased tissue. In addition to the heterogeneity of diseased tissues observed from patient to patient (136), the sampling and processing of biofluids is minimally invasive and requires minimal cost and technical expertise (54). These advantages become of even greater importance when the biomarker is later translated to a clinical setting. Analysis of biofluids is hindered by the lack of a corresponding genome or transcriptome that allows gene expression to be measured (117). Thus, over the last several years an extraordinary amount of effort has

been spent studying the protein component of human bodily fluids in hopes of identifying novel biomarkers of disease. It is necessary to address the characteristics of these fluids, their diagnostic potential, and the obstacles to consider in their investigation.

2.2.1 Plasma/Serum

Plasma is obtained when blood is withdrawn in the presence of an anticoagulant, and centrifuged to remove cellular elements. A serum sample is obtained if the blood is instead allowed to clot before centrifugation. The choice of whether to analyze sera or plasma has not been adequately addressed, yet the protein compositions of these two samples differ significantly (54). The characterization of plasma and serum is led by the Plasma Proteome Project (PPP) organized by the Human Proteome Organization (HUPO), an international scientific organization focused on identifying large numbers of proteins in biofluids (87). The PPP has identified 3020 proteins thus far from plasma and serum through the combined efforts of 18 laboratories (87).

The complexity of the plasma proteome has posed a significant barrier to its analysis. The protein concentration range of plasma proteins spans up to twelve orders of magnitude, while current proteomic approaches are capable of only covering three to four orders of magnitude (8, 71, 136). Plasma has an average protein concentration of 80 mg/ml, yet many common tumor markers are found in the ng/ml range (136). Thus, without first reducing the complexity of plasma, it would be unfeasible to identify most biomarkers with the current MS-based methods. Therefore, prefractionation techniques, further discussed below, are required to achieve a more comprehensive analysis of plasma/serum proteins. However, specific to sera is the presence of seven abundant proteins (albumin, large-chain immunoglobulins (Igs), small-chain Igs, transferrin, α-

macroglobulin, antitrypsin, and haptoglobulin) that account for 97% of all proteins (135), and the 22 most abundant proteins account for 99% of the sera protein content (117). Thus many studies commonly employ commercially available multiple affinity removal columns (MARC) to deplete plasma of the most abundant proteins (54). In addition many groups develop in-house immunoabsorbers for further depletion (136). Although these depletion methods provide a simple means for enriching proteins of low abundance, they may also lead to a loss of diagnostic information. One study found that 210 proteins and peptides were associated in complexes with the very abundant species, and 12 of these are currently utilized as clinical biomarkers (131). The association of a biomarker in a protein complex may be especially relevant to its diagnostic utility; as such complexes are not adequately cleared by the renal system (131).

There has been an increased interest in the diagnostic potential of the low-molecular weight (LMW) range (<50,000 Da) of the serum/plasma proteome, also referred to as the peptidome. This set of molecules may provide specific information about the diseased tissue. Proteins originating from diseased tissue are often too large to diffuse through the endothelium and into circulation, but fragments of the parent molecule may be represented in the plasma (94). Such fragments are thought to have particular relevance in the diagnosis of cancer, as formation of fragment isoforms may be a result of unique events in the affected cancerous tissue. These microenvironments often possess high levels of enzymatic and proteinase activity compared to normal tissue. Therefore, specific fragment isoforms may result from unique cleavage sites in the parent molecule, or may be modified with specific PTMs (94). For instance, one study identified peptide sets that could discriminate between prostate, breast, and bladder

cancer, and analyses of the most significant peptides suggested they possessed characteristic termini resulting from unique cleavage sites (118). Thus, these unique isoforms provided a specific signature of the affected tissue, and may have an even greater diagnostic capability than the parent molecules from which they originated. In addition to truncations, post-translationally-modified peptides often represent especially useful biomarkers, and strategies to enrich for glycopeptides prior to analysis have yielded encouraging results (129). Finally, peptide biomarkers are often bound to carrier-proteins, such as albumin, and thus acquire the longer half-life of such proteins. Therefore, studies have attempted to directly isolate and characterize such plasma complexes in the quest for diagnostic molecules (81).

2.2.2 Urine

Urine is an alternative to plasma for biomarker discovery for many systemic diseases. This biofluid is of considerable less complexity (54), although a combination of proteomic studies have identified at least 1000 different gene products (96). This biofluid has been especially promising in the diagnosis of bladder cancer and renal disease, and is of particular use in the study of drug metabolism (117). There is considerable overlap in the plasma and urinary proteomes, as evidenced by a urinary proteome study that found one third of its identified proteins were also found in plasma (95). Urine is composed of three protein sources: soluble proteins, solid-phase components, and exosomes. Soluble proteins account for 49% of the protein content result from glomerular filtration of plasma proteins and proteins secreted by epithelial cells. The glomerular filter retards passage of high-molecular-weight proteins, and molecules <10 kDa are freely filtered. Solid phase components account for 48% of the protein content and constitute whole

epithelial cells or cell fragments such as the plasma membrane and intracellular components. Exosomes are membrane vesicles that originate as internal vesicles of cells that are excreted into the extracellular space, and these represent 3% of the urinary protein content (96).

Analysis of urine requires special considerations. Urinary protein concentrations (~10 μg/ml) are much lower than plasma protein concentrations (~80 mg/ml), and urine contains high levels of salts that can interfere with downstream analytical procedures; therefore, desalting and concentration of samples is required (65). The most abundant protein in urine is the Tamm-Horsfall protein (uromodulin), and this protein forms networks of fibers that can interfere with downstream biomolecular separation techniques. To combat fiber formation, depolymerization can be achieved by the addition of reducing agents and warming (96). Similar to plasma, several simple prefractionation techniques are available to reduce the protein complexity of urine. Depending on the anticipated origin of the biomarker, differential centrifugation can easily lead to isolation of soluble proteins, solid-phase components, and exosomes (96). Albumin, α-1-acid glycoprotein, and IgG are abundant proteins in urine, and can be immunodepleted with MARC (95).

2.2.3 Specialized biofluids

The majority of biofluid proteome studies have focused on plasma/serum and urine. However, analysis of other biofluids can provide more specialized information regarding the local diseased tissue from which they sample. For instance, cerebrospinal fluid (CSF), a clear liquid that protects the central nervous system from trauma, is in direct contact with the extracellular space of the brain, and the protein profile of this

biofluid may reflect physiological processes such as aging and the presence of neurodegenerative disease (117). Other biofluids that have undergone proteomic characterization include saliva, bronchoalveolar lavage fluid, synovial fluid, nipple aspirate fluid, tear fluid, and amniotic fluid (54). Generally, the sets of proteins identified in these various biofluids display minimal overlap. However, common to all biofluids are large number of proteins with very wide dynamic ranges in concentration (54).

2.3 Proteomic strategies

2.3.1 Mass spectrometry

MS is the principal technology used for biomarker discovery. The coupling of gas chromatography with MS has been used in the analyses of small molecules for decades. In more recent years the development of two new methodologies to volatize and ionize large molecules, ESI and matrix-assisted laser desorption/ionization (MALDI), has greatly accelerated the field of proteomics (32). In addition to these two types of ion sources, numerous types of mass analyzers continue to be developed with differing sensitivities, mass accuracies, resolution, and scan rates (4, 127). Regardless of the type of instrumentation used, mass spectra are produced which plot the mass-to-charge ratio (m/z) of the ions observed (x-axis) versus the abundance of the ions detected (y-axis). Protein identifications are established through accurate determination of the peptide mass, most often achieved with MALDI-MS or through MS/MS that includes collision-induced dissociation (CID) of precursor ions to generate fragment ion spectra (4). In both cases, spectra are compared with calculated theoretical spectra derived from comprehensive protein sequence databases, and statistical algorithms are used to identify the best matches.

Shotgun-based proteomics combines protein prefractionation with MS to substantially increase the number of proteins identified in a given sample, and this technique is able to partially overcome the vast dynamic range of proteins observed in biofluids. LC-MS/MS coupled with ESI is the most commonly used MS platform for shotgun-proteomics. (125) To further reduce the proteome complexity, fractionation before LC-MS/MS is often achieved with either gel-based or gel-free methods. The gel-based methods have included both 1D-SDS-PAGE and 2D-SDS-PAGE. 2D-SDS-PAGE has been employed for over 25 years and is still a very commonly used fractionation methodology; however, in recent years its popularity for the purpose of proteome characterization has been waning (4). This is partly due to its somewhat limited dynamic range, as demonstrated through a systematic study of the yeast *Saccharomyces cerevisiae* where only the most abundant proteins were routinely identified with this methodology (46). Capillary electrophoresis (CE) is a non-gel based method that can also separate proteins and peptides based on size and pI, and can be directly coupled to a mass spectrometer (58). However, the most common type of non-gel based method is multidimensional protein identification technology (MudPIT) (123) which most commonly separates peptides with strong-cation-exchange chromatography (SCX) followed by reverse-phase chromatography (RP). These steps are physically linked, and the effluent is directly fed into a mass spectrometer. Although this is the dominant separations strategy used in shotgun proteomics, experimental design can vary greatly and continues to expand with increasingly available numbers of chromatography and mass spectrometer combinations (38). As demonstrated with studies on the plasma proteome, the use of alternative discovery platforms often generates very different sets of

protein lists with minimal overlap, suggesting a single shotgun-based proteomics experiment falls short of being comprehensive (13). Although the object of shotgun-based proteomics is usually not biomarker identification, the resulting data contributes greatly to the understanding of the repertoire of proteins present in each biofluid, and this information can help in planning targeted experiments for biomarker discovery.

The most widely-used MS-based platform for biomarker discovery in biofluids is surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) MS technology (ProteinChip®, BioRad) (107). The pioneering work of Petricoin *et al.* demonstrated that this approach could distinguish ovarian cancer from healthy patients with outstanding accuracy (93). Since this seminal study the technology has gained enormous popularity, and a recent of review of serum biomarkers associated with cancer showed that 46 of 98 publications employed SELDI-TOF-MS (54). This technology begins with complex samples, often without prior enzymatic digestion, that are incubated on chips containing eight or 16 spots coated with hydrophobic, ion-exchange, metal affinity, normal phase, or more specialized surfaces (32). Unbound components are washed away and the proteins and peptides are ionized and analyzed in a fashion similar to MALDI-TOF-MS. The resulting peak patterns and peak intensities are compared across control and sample populations, and these data are suggested to reflect peptide concentrations in the biofluid. Sophisticated computer algorithms based on "training" principles are used to identify sets of discriminating peaks that have been suggested to be diagnostic of disease with near 100% accuracy; however, the identity of these peaks are often not determined (32). SELDI-TOF-MS has been used for the analysis of at least 14 types of cancer (54), and

other types of diseases including, but not limited to, Alzheimer's disease (19), atherosclerosis (27), and Creutzfeldt-Jacob disease (45).

2D-SDS-PAGE combined with MS identification has demonstrated considerable success in the identification of biomarkers. In the simplest approach, 2D-SDS-PAGE gels of protein extracts from normal and diseased individuals are stained and compared with densitrometry (58). Improving on this method is the development of difference gel electrophoresis (DIGE). In this approach two samples (e.g. normal vs. diseased) are labelled with two different fluorescent dyes, mixed, and analyzed on the same gel. Scanning the gel at different excitation wavelengths often reveals distinct protein expression patterns, and these aid in identifying proteins that increase in quantity in the diseased state. (58) Both the standard 2D-SDS-PAGE comparison and DIGE approaches have been used for biomarker discovery in numerous types of cancer (58), and in such diverse diseases as Alzheimer's disease (76), cystic fibrosis (102), amyotrophic lateral sclerosis (42), arthritis (108, 116), Parkinson's disease (42), liver disease (5), and severe acute respiratory syndrome (SARS) (21). Finally, analyses of tissue and blood protein extracts by 2D-SDS-PAGE and western blotting with sera can help in the identification of autoantibodies (58). Such antibodies can be indicative of autoimmune disease (56). Additionally, autoantibodies are often raised against tumor proteins, and identification of their presence with the Western blot approach has been shown to have potential for the diagnosis of cancer (69). The analyses of pathogen protein extracts with sera has seen widespread use for the determination of infection, and successful examples include TB infection (103, 104) and fungal infection (97). Again, the main disadvantages of all of these gel-based methods are limited sensitivity and throughput.

Studies have attempted to demonstrate the ability of LC-MS to accurately quantify peptide peaks in complex samples (72), and software has been specifically designed for this purpose (75). However, the prevailing view is that "the relationship between the amount of analyte present and measured signal intensity is complex and incompletely understood", and ..."mass spectrometers are therefore inherently poor quantitative devices" (cited in ref 4). Thus, labeling strategies including labeled proteolysis, isotope-coded affinity tag (ICAT) technology, and "isotope tags for relative and absolute quantitation" (iTRAQ™, Applied Biosystems) technology have been developed to enable MS-based quantitation for biofluid-derived biomarkers. In each of these cases the samples are differentially labeled prior to MS analyses, and the ratio of the peaks eluting from the LC column are compared in order to determine the relative abundance of the protein in each sample. Identities are usually determined via MS/MS (96). The iTRAQ™ technology has been used to identify potential biomarkers from biofluids that are indicative of endometrial carcinoma (29) and Alzheimer's disease (2), and ICAT has been used to discover biofluid-derived biomarkers of breast cancer (89), pancreatic cancer (22), and brain injury (50).

Following the biomarker discovery phase, the ability of these candidate molecules to distinguish between disease states should be independently validated. Although the biomarker discovery phase has received more attention, some believe the validation phase likely poses an even greater challenge (136). Validation is routinely achieved through standard immunoassays (ELISA), or to a lesser extent, targeted antibody arrays (discussed below). Recently an MS-based method termed multiple reaction monitoring (MRM) has been developed for biomarker validation (66). This approach requires

knowledge of the precursor mass and the fragmentation spectra specific to an analyte, and programming the mass spectrometer to filter data with these two mass identities provides a specific assay. Furthermore, these assays depend on internal spiked standards for accurate quantitation, and multiplexing allows for the simultaeneous analysis of 30 or more analytes in one MS run. (6) Although this technique is not yet in widespread use, there is a great deal of support for this application. (119)

2.3.3 Protein microarrays

The fundamental principles behind microarray technology were described by Roger Ekins *et al.* almost two decades ago. This work developed the "ambient analyte theory" that explained why microspot assays could achieve greater sensitivity than standard ligand-binding assays (35-37). Putting this theory to the test, rapid progress in genome-sequencing efforts resulted in the fabrication of single slides that could accommodate tens of thousands of oligonucelotides, and these arrays provide for the ability to monitor gene-expression on a global scale (105). Protein microarray technology has directly benefited from advances in DNA microarray technology, and uses the same instrumentation, automation, detection reagents, and even bioinformatic software (114). Instead of oligonucleotides as the ligands, protein microarrays consist of immobilized proteins, peptides, antibodies, or aptamers. The applications of protein microarrays are very diverse and include the evaluation of protein-protein interactions (132), protein-nucleic acid interactions (47), protein-lipid interactions (132), protein-drug interactions (55), protein function (134), protein-receptor interactions (60), identification of PTMs (110), and determination of antibody specificity (82). In addition, microarrays can be used for both discovery and validation of disease biomarkers.

The antibody microarray consists of specific antibodies that are immobilized and used to probe complex biofluids such as serum. Similar to a capture ELISA, this high-throughput format provides the advantages of miniaturization, parallelism, and automation. This format has successfully discriminated between healthy sera and those with prostate cancer (83), pancreatic cancer (88), and lung cancer (88). Due to the nature of the assay, knowledge of the antibody specificity allows for direction determination of the identities of the overexpressed proteins. Yet this can also be a disadvantage, because the antibody microarray is always limited by the number and quality of the antibody panel.

Microarrays can also be fashioned in a similar manner as the Western blot, where proteins are immobilized and antibodies in the sera are detected. Advantages over the Western blot include the ability to accommodate thousands of individual antigens or antigen pools on a single slide, the presentation of antigens in a more native state and of equal abundance, and the ability to generate hundreds of slides at one time leading to increased reproducibility. These antigen arrays have been used to detect antibodies in a variety of autoimmune diseases (61), including lupus (74), alopecia areata (77), and arthritis (78). Antigen arrays have also been used to assess the presence of Type I allergies (52). Using protein fractions derived from tumor cell lines, an antigen array was able to distinguish healthy sera from prostate cancer sera (16), and lung cancer sera (99). These types of autoantibody studies can now take advantage of commercially available high-density protein microarrays (ProtoArray®, Invitrogen) that represent greater than 5000 human proteins.

Antigen arrays have also been extended to infectious disease research. In an initial proof-of principle study, Bacarese-Hamilton *et al.* (10) immobilized recombinant antigens of various pathogens to glass slides, and human antibodies specific for each antigen were detected in sub-picogram amounts. This microarray assay also performed at the same level of efficiency as conventional ELISA-based methods in differentiating between positive and negative sera. Following this initial study, larger microarray panels of recombinant proteins have successfully differentiated infected versus noninfected sera for pathogens such as *Neisseria meningitidis* (111), *M. leprae* (44) *Plasmodium falciparum* (43), Hepatitis B and C viruses (34), and the SARS coronavirus (133). Furthermore, microarray serum profiling has been extended to characterize antibody responses generated upon vaccination with *Yersinia pestis* (73), *Francisella tularensis* (113), vaccinia virus (26), and simian HIV (86). In addition to analyzing simple antigen-antibody interactions, more complex peptide-MHC (pMHC) arrays are being developed that are able to detect the cellular response (30, 109, 112). These advanced arrays detect cytokine secretion from T cells that bind immobilized pMHCs and costimulatory molecules, and the results often reveal informative epitopes that could be utilized for diagnosis or vaccine design (20). Finally, arrays of immobilized host glycoprotein receptors were suggested to be capable of detecting whole pathogens from complex mixtures, including serum (33), although improvements will have to be made to achieve the sensitivity needed for clinical use.

In addition to biofluids, tissues and cellular lysates can also be analyzed for biomarker discovery. This is the basis behind the newly developed reverse phase (RP) array. RP arrays consist of panels of crude cellular lysates or immobilized tissue samples

from many different patients that are obtained through laser capture microdissection. Probing the tissues with specific antibodies against various signaling pathway molecules or PTMs can lead to deeper insights into the biology of the diseased tissue. Additionally, this research should lead to the identification of biomarkers that differentiate healthy from diseased patient samples, and has specifically been shown to be useful in the prognosis of breast cancer (79).

2.4 Progress, challenges, and criticisms

Currently a very small number of cancer biomarkers are routinely used in the clinical setting. These "classical" markers include α-Fetoprotein, CEA, prostate-specific antigen (PSA), CA125, CA15.3, CA19.9, Igs, hCG, and steroid hormone receptors. These biomarkers are mostly recommended for the monitoring of therapy and treatment response. Only PSA, a molecule specific for prostate cancer, has been approved for population screening by the Food and Drug Administration. Many of these biomarkers were initially discovered through development of monoclonal antibodies that recognize "cancer-associated antigens", identified through immunization of animals with tumor extracts or cancer cell lines. (32) In contrast, the proteomics process of biomarkers discovery has led to very few, if any, novel biomarkers specific for either cancer or other types of disease that have resulted in a commercial product approved by regulatory agencies (136). Thus the intense excitement regarding the strategies described in section 2.3 has not been matched with the development of real-world clinical tools. One possible reason for this disparity is that the current system does not easily allow for the transition of biomarkers from the discovery phase to the validation phase. Drug development is mostly led by the pharmaceutical industry, where a single pipeline takes drugs from the

discovery phase all the way to clinical trials. In contrast, diagnostic biomarker discovery is often driven by academic and government funded research, while commercial diagnostics perform clinical validation. (7) In addition to this gap in research infrastructure, numerous concerns have been raised regarding the potential for proteomics to accurately identify biomarkers of disease. These concerns question numerous aspects regarding the technical limitations of MS, experimental design, statistical analyses, and the meaning of the data obtained (119).

Despite advances in proteomics, the number of new protein biomarkers approved by regulatory agencies has actually declined in the last decade (136). Many believe that the most powerful single cancer biomarkers have already been identified. (32) Thus, there is little hope that a single biomarker, or "golden bullet", has the ability to describe complex diseases (136). However, there is great optimism (98) that discovery of other biofluid molecules will augment the capabilities of the current biomarkers, and that diagnostic algorithms could identify combinations that provide improved sensitivity and specificity as a whole. The majority of the data supportive of this hypothesis has been generated by SELDI-TOF-MS technology. However, critics have pointed out several shortcomings of this methodology and suggest some data has been misinterpreted. (32) For instance, the classical cancer biomarkers, including PSA, have rarely been identified with SELDI-TOF-MS. These biomarkers are in the 2-150 pM range (126), while the newly discovered putative cancer biomarkers are typically in the 1-40 μM range (24, 53, 128, 130). Thus, SELDI-TOF-MS is likely biased towards molecules of high abundance. This is of particular relevance, because many of the highly abundant proteins found in plasma are derived from by the liver, and often represent acute-phase proteins that are

indicative of inflammation. Such proteins have long been known to be unspecific for the presence of cancer, and for this reason their clinical use has been avoided. Adding to this problem, a significant portion of SELDI-TOF-MS studies publish only discriminating peak values without providing peptide identities, increasing the potential for these peaks to represent nonspecific molecules. (32) Furthermore, the abundance of specific molecules that are truly associated with cancer is likely to be very low. If one considers that a typical tumor represents approximately 0.006% of the total body volume (117), the ability of a small localized tumor to modify the serum proteomic pattern is very limited.

Another concern regarding biomarker proteomics results is the large data discrepancy reported across different research groups investigating the same phenomena (32). As previously discussed, the sheer complexity of biofluids often generates data sets with minimal overlap, and this may partially explain the reported disparities (13). However, even reports that use the same SELDI-TOF-MS surface chemistry (IMAC-Cu) to study the same biofluid (sera) for the diagnosis of the same disease (prostate cancer) identified very different sets of discriminating peaks (3, 12, 32, 70). Two of these studies showed that the use of differing bioinformatic algorithms produced very different sets of data (3, 12), and one study demonstrated that "noise" peaks can perfectly discriminate between normal and cancer patients when training algorithms were used (11). In addition to data analyses, numerous factors could contribute to these discrepancies such as the physiological status of the patient cohort used or more subtle variations in experimental technique. There has even been described *ex vivo* exoprotease activity in sera that generates truncated peptides when samples are handled differently (80, 118). These reasons suggest that steps should be taken to standardize comparisons within and across

laboratories. Specifically, spiking samples with an internal protein reference standard should help in comparing sera samples (136). Another specific example is that reproducibility of the urine sample analyses can be affected by alterations in protein concentrations due to water content. Therefore, normalizing data to creatinine excretion rates has been suggested (96). The standardization of such biofluid analytical procedures is one of the goals of the HUPO PPP (87). Finally, identification of the discriminating peaks at the amino acid level would help to determine whether different laboratories observe the same proteins, and aid in the independent validation of newly identified biomarkers. This approach could also target those biomarkers that are likely to be specific for the disease of interest, establishing a cause and effect relationship. (32)

Biomarker discovery should ultimately lead to the development of applied clinical assays. A desirable diagnostic assay should be of low cost, be minimally invasive, require minimal sample preparation, employ easy-to-operate technology, be fast and easy to interpret, and provide 100% sensitivity and specificity (58). Although it is not inconceivable that mass spectrometers and proteins microarrays could be used in the clinical laboratory, the prevailing view is that ELISA-based formats are the best available option to achieve these criteria (101). Therefore, knowledge of the biomarker identity will be required for development of antibodies used in the ELISA (32). This change in format from MS-based methods to ELISAs requires special consideration (120). For instance, an ELISA is capable of processing only a very limited panel of biomarkers; therefore, very complex discriminating patterns determined through MS-based discovery may be of little utility (7).

In summary, proteomics in the context of biomarker discovery is a hotly debated (31, 92) and rapidly evolving field. Perhaps success will be soon be demonstrated as the results of ongoing clinical trials are revealed (32). Furthermore, regardless of the current status future advances in technology, such as increased separation efficiency and MS resolution, should provide optimism.

2.5 TB biomarkers

Only a handful of approaches have applied proteomic technologies to discover TB biomarkers (25, 67, 103, 124). Nevertheless, there has been an intense research interest in developing techniques to determine *Mtb* infection or to assess responses to drug treatment. This section briefly addresses the currently used clinical and micobiological tests, and further examines molecular biomarkers that have been identified to date.

2.5.1 Markers of infection

Few diagnostic approaches are able to differentiate TB disease from *Mtb* infection, and most diagnostic approaches usually assess one or the other. For instance, a chest X-ray can indicate general pulmonary disease, however, this clinical symptom may not be caused by *Mtb*. Alternative, the PPD skin test indicates previous exposure to *Mycobacterium* spp., but does not describe TB disease. As mentioned in section 1.4, a goal of TB immunodiagnostics is to not only establish *Mtb* infection but also determine the disease state.

Perhaps the most fundamental marker of *Mtb* infection is direct observation of the bacillus. As described in section 1.4 detection is often achieved through identification of AFB in the sputum with microscopy. More sensitive detection of *Mtb* is achieved through cultivation on solid media, which may take several weeks for visible detection of

the organism. Improvements on this method have been made which provide more rapid detection in broth-based cultures, including the BACTEC 460 system and the Mycobacteria Growth Indicator Tube (MGIT), both of which measure metabolic activity. Confirmation of *Mtb* isolation is routinely achieved with biochemical tests or nucleic acid probe assays (Gen-Probe). (49)

Section 1.4 addressed mycobacterial antigens that are indicative of *Mtb* infection. The biomarkers in these cases are either molecular in nature (specific antibodies detected in the sera) or phenotypic in nature (cellular responses). In addition to evaluating the host immune response, several studies have attempted to directly identify mycobacterial products from biofluids. An advantage of this approach is that an intact host immune response is not required for detection, which may facilitate diagnosis of immunocompromised (including HIV+TB+) individuals. Furthermore, the release of mycobacterial molecules must precede their recognition by the immune system; which could, in theory, lead to earlier detection. Detection is usually achieved via capture ELISA using antibodies developed specifically for the molecule of interest. For pulmonary TB patients, LAM was identified in both sputum (90) and urine (15, 48, 115), the Antigen 85 complex was identified in sera (14, 63), and the 38-kDa antigen was found in sera (68). Furthermore, Ag85B and Ag85C were found in the CSF of tuberculous meningitis patients (62). Finally, *Mtb* DNA can be detected in clinical specimens, and is routinely identified through commercially available nucleic acid amplifications assays such as MTD and Amplicor (106).

2.5.2 Markers of drug treatment response

In addition to biomarkers that indicate *Mtb* infection, there has been an increasing interest in identifying both host and mycobacterial biomarkers that can forecast the outcome of a particular drug treatment. Such biomarkers would be especially useful in clinical trials to determine whether a novel drug regimen is superior to standard existing drug regimens. This type of biomarker, termed a "surrogate endpoint", would ideally be indicative of the drug's bactericidal activity and sterilizing activity (91). Sterilization is an especially important TB drug attribute, as the persistence of bacteria with altered physiological states in different anatomical sites often leads to reactivation (relapse). However, definitive determination of relapse rates can only be achieved through life-long monitoring of the patients. Instead, many clinical trials define cured patients as those who consistently culture negative over a follow-up period of 12 to 24 months after the end of treatment (39). Studies that attempt to further address relapse rates are hindered by the very large number of patients and long follow-up periods required to complete the study (91). Thus biomarkers that can predict drug efficacy early in clinical trials would be of great utility.

Currently drug treatment responses are evaluated based on the detection of bacilli through a variety of different manners. The WHO definition of cure is sputum smear negativity in the last month of treatment and on at least one previous occasion (1). However, this method would have several disadvantages in clinical trials, including low sensitivity, high variability, detection of non-viable organisms, and the fact more than half of patients with pulmonary TB are smear-negative at diagnosis (91). Instead, clinical trials most commonly evaluate the early bactericidal activity of drugs by measuring the

decrease in colony forming units (CFU) in sputum over a two to seven day period after initiation of treatment (91). Although CFU counts continue to drop over more extended periods, the initial CFU fall was found to be more indicative of bactericidal activity (59). Pitfalls of this method include high variability intrinsic to counting CFUs from sputum, the inability to assess responses in extrapulmonary TB patients, and the inability to assess a drug's sterilizing activity. However, extending CFU counts longer than 28 days has been suggested to overcome the latter pitfall (17, 41). Assessing the number of patients' sputum samples that become culture positive two months post-treatment is the only accepted method of determining sterilizing activity (91), as extended follow-up studies showed a positive correlation between two-month sputum culture conversion and relapse rates (84). Expanding on this strategy, determination of the amount of time it takes to turn culture-negative may be even more informative, and this method is under current investigation (91).

Direct detection of mycobacterial molecular biomarkers would provide a much more rapid measurement of drug treatment response compared to culture. Several studies have attempted to correlate concentrations of these molecular biomarkers in sputum with treatment responses. Mycobacterial DNA and ribosomal RNA levels in sputum samples were shown to fall with effective drug therapy (64, 85). However, these molecules remained detectable long after the conversion of sputum smears and culture results to negative, suggesting this methodology may not distinguish viable from non-viable organisms. In contrast, the more labile nature of messenger RNA (mRNA) may be advantageous. Antigen 85B mRNA levels in the sputum cleared very rapidly and corresponded with the early loss of viable organisms (28, 51). In addition to nucleic acid

detection, Antigen 85 complex protein was detected in patient sputum, and increased protein concentrations at day 14 of treatment were found only in patients whom disease persisted beyond 90 days, suggesting this protein may have the potential to predict relapse (121). This predictive capability may be specific to the Antigen 85 complex, as the expression of this complex is induced with isoniazid, and the complex is thought to play a role in bacterial nonreplicating persistence (40, 121).

Upon treatment with effective drugs, various physiological features of the host are expected to change and could be used as markers of treatment response. Chest radiography is often indicative of bacterial load and disease extent; however, radiographs improve more slowly than the clinical response, thus limiting their use as an early indicator (91). The immune response has been shown to change with clinical improvement, and to correlate with effective treatment. Specifically, T cell responses to ESAT-6 peptides were shown to drop significantly in patients that responded to three months of a standard drug regimen (18). Furthermore, cytokine levels in the sputum were shown to drop following four weeks of drug treatment and these levels were associated with mycobacterial clearance (100). Recently a whole blood bactericidal assay was developed, where blood is mixed with *Mtb* and the rate of bacterial killing, via immune mechanisms, is measured (23). Bactericidal activity was generally greater in patient's blood that responded better to treatment, suggesting this assay could be used a marker of patient prognosis (122). In terms of the serological response, antibody titers generally increased while antibody avidities generally decreased during drug treatment; however, large variations among individual serological responses may limit their utility (9, 100). Host molecular markers have been shown to change with drug treatment.

Acute-phase proteins, discussed in section 2.4, were shown to generally decrease in concentration in serum samples when patients responded well to drug treatment (57). However, these protein concentrations were also variably abnormally high among patients at the beginning of treatment, resulting in low sensitivity for these markers. In addition, these inflammatory molecules are likely to provide low specificity for *Mtb* cure.

2.6 Rationale and objectives

Recent years have witnessed significant technological advances for the identification and evaluation of biomarkers of disease. However, this progress has largely been driven by efforts to establish biomarkers indicative of cancer, and have not transitioned well to infectious disease. Worldwide TB control efforts would be greatly aided by the development of a robust diagnostic assay that is based on host or bacterial biomarkers. To date the discovery and evaluation of TB serodiagnostic markers has been achieved through experimental approaches that have been very focused in nature. In order to continue to move this field forward, more global strategies are needed to further define the repertoire of antigens recognized by the immune system. This work applies emerging proteomic technologies to identify and characterize TB serodiagnostic antigens. Furthermore, this work tests the hypothesis that a defined antibody response against a subset of antigens is generated that is dependent on the state of TB disease progression. The results should contribute to the rational design of a successful future serodiagnostic assay.

2.7 Literature cited

1. 2001. Revised international definitions in tuberculosis control. Int J Tuberc Lung Dis **5**(3): 213-5.

2. **Abdi, F., J. F. Quinn, J. Jankovic, M. McIntosh, J. B. Leverenz, E. Peskind, R. Nixon, J. Nutt, K. Chung, C. Zabetian, A. Samii, M. Lin, S. Hattan, C. Pan, Y. Wang, J. Jin, D. Zhu, G. J. Li, Y. Liu, D. Waichunas, T. J. Montine, and J. Zhang**. 2006. Detection of biomarkers with a multiplex quantitative proteomic platform in cerebrospinal fluid of patients with neurodegenerative disorders. J Alzheimers Dis **9**(3): 293-348.

3. **Adam, B. L., Y. Qu, J. W. Davis, M. D. Ward, M. A. Clements, L. H. Cazares, O. J. Semmes, P. F. Schellhammer, Y. Yasui, Z. Feng, and G. L. Wright, Jr.** 2002. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res **62**(13): 3609-14.

4. **Aebersold, R. and M. Mann**. 2003. Mass spectrometry-based proteomics. Nature **422**(6928): 198-207.

5. **Amacher, D. E., R. Adler, A. Herath, and R. R. Townsend**. 2005. Use of proteomic methods to identify serum biomarkers associated with rat liver toxicity or hypertrophy. Clin Chem **51**(10): 1796-803.

6. **Anderson, L. and C. L. Hunter**. 2006. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol Cell Proteomics **5**(4): 573-88.

7. **Anderson, N. L.** 2005. The roles of multiple proteomic platforms in a pipeline for new diagnostics. Mol Cell Proteomics **4**(10): 1441-4.

8. **Anderson, N. L. and N. G. Anderson**. 2002. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics **1**(11): 845-67.

9. **Arias-Bouda, L. M., S. Kuijper, A. Van der Werf, L. N. Nguyen, H. M. Jansen, and A. H. Kolk**. 2003. Changes in avidity and level of immunoglobulin G antibodies to *Mycobacterium tuberculosis* in sera of patients undergoing treatment for pulmonary tuberculosis. Clin Diagn Lab Immunol **10**(4): 702-9.

10. **Bacarese-Hamilton, T., F. Bistoni, and A. Crisanti**. 2002. Protein microarrays: from serodiagnosis to whole proteome scale analysis of the immune response against pathogenic microorganisms. Biotechniques **Suppl**: 24-9.

11. **Baggerly, K. A., J. S. Morris, and K. R. Coombes**. 2004. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics **20**(5): 777-85.

12. **Banez, L. L., P. Prasanna, L. Sun, A. Ali, Z. Zou, B. L. Adam, D. G. McLeod, J. W. Moul, and S. Srivastava**. 2003. Diagnostic potential of serum proteomic patterns in prostate cancer. J Urol **170**(2 Pt 1): 442-6.

13. **Barnea, E., R. Sorkin, T. Ziv, I. Beer, and A. Admon**. 2005. Evaluation of prefractionation methods as a preparatory step for multidimensional based chromatography of serum proteins. Proteomics **5**(13): 3367-75.

14. **Bentley-Hibbert, S. I., X. Quan, T. Newman, K. Huygen, and H. P. Godfrey**. 1999. Pathophysiology of antigen 85 in patients with active tuberculosis: antigen 85 circulates as complexes with fibronectin and immunoglobulin G. Infect Immun **67**(2): 581-8.

15. **Boehme, C., E. Molokova, F. Minja, S. Geis, T. Loscher, L. Maboko, V. Koulchin, and M. Hoelscher**. 2005. Detection of mycobacterial lipoarabinomannan with an antigen-capture ELISA in unprocessed urine of Tanzanian patients with suspected tuberculosis. Trans R Soc Trop Med Hyg **99**(12): 893-900.

16. **Bouwman, K., J. Qiu, H. Zhou, M. Schotanus, L. A. Mangold, R. Vogt, E. Erlandson, J. Trenkle, A. W. Partin, D. Misek, G. S. Omenn, B. B. Haab, and S. Hanash**. 2003. Microarrays of tumor cell derived proteins uncover a distinct pattern of prostate cancer serum immunoreactivity. Proteomics **3**(11): 2200-7.

17. **Brindle, R., J. Odhiambo, and D. Mitchison**. 2001. Serial counts of *Mycobacterium tuberculosis* in sputum as surrogate markers of the sterilising activity of rifampicin and pyrazinamide in treating pulmonary tuberculosis. BMC Pulm Med **1**: 2.

18. **Carrara, S., D. Vincenti, N. Petrosillo, M. Amicosante, E. Girardi, and D. Goletti**. 2004. Use of a T cell-based assay for monitoring efficacy of antituberculosis therapy. Clin Infect Dis **38**(5): 754-6.

19. **Carrette, O., I. Demalte, A. Scherl, O. Yalkinoglu, G. Corthals, P. Burkhard, D. F. Hochstrasser, and J. C. Sanchez**. 2003. A panel of cerebrospinal fluid potential biomarkers for the diagnosis of Alzheimer's disease. Proteomics **3**(8): 1486-94.

20. **Chen, D. S. and M. M. Davis**. 2006. Molecular and functional analysis using live cell microarrays. Curr Opin Chem Biol **10**(1): 28-34.

21. **Chen, J. H., Y. W. Chang, C. W. Yao, T. S. Chiueh, S. C. Huang, K. Y. Chien, A. Chen, F. Y. Chang, C. H. Wong, and Y. J. Chen**. 2004. Plasma proteome of severe acute respiratory syndrome analyzed by two-dimensional gel electrophoresis and mass spectrometry. Proc Natl Acad Sci U S A **101**(49): 17039-44.

22. **Chen, R., S. Pan, E. C. Yi, S. Donohoe, M. P. Bronner, J. D. Potter, D. R. Goodlett, R. Aebersold, and T. A. Brentnall**. 2006. Quantitative proteomic profiling of pancreatic cancer juice. Proteomics **6**(13): 3871-9.

23. **Cheon, S. H., B. Kampmann, A. G. Hise, M. Phillips, H. Y. Song, K. Landen, Q. Li, R. Larkin, J. J. Ellner, R. F. Silver, D. F. Hoft, and R. S. Wallis**. 2002. Bactericidal activity in whole blood as a potential surrogate marker of immunity after vaccination against tuberculosis. Clin Diagn Lab Immunol **9**(4): 901-7.

24. **Cho, W. C., T. T. Yip, C. Yip, V. Yip, V. Thulasiraman, R. K. Ngan, T. T. Yip, W. H. Lau, J. S. Au, S. C. Law, W. W. Cheng, V. W. Ma, and C. K. Lim**. 2004. Identification of serum amyloid a protein as a potentially useful biomarker to monitor relapse of nasopharyngeal cancer by serum proteomic profiling. Clin Cancer Res **10**(1 Pt 1): 43-52.

25. **Covert, B. A., J. S. Spencer, I. M. Orme, and J. T. Belisle**. 2001. The application of proteomics in defining the T cell antigens of *Mycobacterium tuberculosis*. Proteomics **1**(4): 574-86.

26. **Davies, D. H., X. Liang, J. E. Hernandez, A. Randall, S. Hirst, Y. Mu, K. M. Romero, T. T. Nguyen, M. Kalantari-Dehaghi, S. Crotty, P. Baldi, L. P. Villarreal, and P. L. Felgner**. 2005. Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery. Proc Natl Acad Sci U S A **102**(3): 547-52.

27. **Dayal, B. and N. H. Ertel**. 2002. ProteinChip technology: a new and facile method for the identification and measurement of high-density lipoproteins apoA-I and apoA-II and their glycosylated products in patients with diabetes and cardiovascular disease. J Proteome Res **1**(4): 375-80.

28. **Desjardin, L. E., M. D. Perkins, K. Wolski, S. Haun, L. Teixeira, Y. Chen, J. L. Johnson, J. J. Ellner, R. Dietze, J. Bates, M. D. Cave, and K. D. Eisenach**. 1999. Measurement of sputum *Mycobacterium tuberculosis* messenger RNA as a surrogate for response to chemotherapy. Am J Respir Crit Care Med **160**(1): 203-10.

29. **DeSouza, L. V., J. Grigull, S. Ghanny, V. Dube, A. D. Romaschin, T. J. Colgan, and K. W. Siu**. 2007. Endometrial carcinoma biomarker discovery and verification using differentially tagged clinical samples with multidimensional

liquid chromatography and tandem mass spectrometry. Mol Cell Proteomics **6**(7): 1170-82.

30. **Deviren, G., K. Gupta, M. E. Paulaitis, and J. P. Schneck**. 2007. Detection of antigen-specific T cells on p/MHC microarrays. J Mol Recognit **20**(1): 32-8.

31. **Diamandis, E. P.** 2003. Point: Proteomic patterns in biological fluids: do they represent the future of cancer diagnostics? Clin Chem **49**(8): 1272-5.

32. **Diamandis, E. P.** 2004. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. Mol Cell Proteomics **3**(4): 367-78.

33. **Disney, M. D. and P. H. Seeberger**. 2004. The use of carbohydrate microarrays to study carbohydrate-cell interactions and to detect pathogens. Chem Biol **11**(12): 1701-7.

34. **Duan, L., Y. Wang, S. S. Li, Z. Wan, and J. Zhai**. 2005. Rapid and simultaneous detection of human hepatitis B virus and hepatitis C virus antibodies based on a protein chip assay using nano-gold immunological amplification and silver staining method. BMC Infect Dis **5**: 53.

35. **Ekins, R. and F. Chu**. 1992. Multianalyte microspot immunoassay. The microanalytical 'compact disk' of the future. Ann Biol Clin (Paris) **50**(5): 337-53.

36. **Ekins, R., F. Chu, and E. Biggart**. 1990. Multispot, multianalyte, immunoassay. Ann Biol Clin (Paris) **48**(9): 655-66.

37. **Ekins, R. P.** 1989. Multi-analyte immunoassay. J Pharm Biomed Anal **7**(2): 155-68.

38. **Fournier, M. L., J. M. Gilmore, S. A. Martin-Brown, and M. P. Washburn**. 2007. Multidimensional separations-based shotgun proteomics. Chem Rev **107**(8): 3654-86.

39. **Fox, W., G. A. Ellard, and D. A. Mitchison**. 1999. Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946-1986, with relevant subsequent publications. Int J Tuberc Lung Dis **3**(10 Suppl 2): S231-79.

40. **Garbe, T. R., N. S. Hibler, and V. Deretic**. 1996. Isoniazid induces expression of the antigen 85 complex in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother **40**(7): 1754-6.

41. **Gillespie, S. H. and B. M. Charalambous**. 2003. A novel method for evaluating the antimicrobial activity of tuberculosis treatment regimens. Int J Tuberc Lung Dis 7(7): 684-9.

42. **Goldknopf, I. L., E. A. Sheta, J. Bryson, B. Folsom, C. Wilson, J. Duty, A. A. Yen, and S. H. Appel**. 2006. Complement C3c and related protein biomarkers in amyotrophic lateral sclerosis and Parkinson's disease. Biochem Biophys Res Commun **342**(4): 1034-9.

43. **Gray, J. C., P. H. Corran, E. Mangia, M. W. Gaunt, Q. Li, K. K. Tetteh, S. D. Polley, D. J. Conway, A. A. Holder, T. Bacarese-Hamilton, E. M. Riley, and A. Crisanti**. 2007. Profiling the antibody immune response against blood stage malaria vaccine candidates. Clin Chem **53**(7): 1244-53.

44. **Groathouse, N. A., A. Amin, M. A. Marques, J. S. Spencer, R. Gelber, D. L. Knudson, J. T. Belisle, P. J. Brennan, and R. A. Slayden**. 2006. Use of protein microarrays to define the humoral immune response in leprosy patients and identification of disease-state-specific antigenic profiles. Infect Immun **74**(11): 6458-66.

45. **Guillaume, E., C. Zimmermann, P. R. Burkhard, D. F. Hochstrasser, and J. C. Sanchez**. 2003. A potential cerebrospinal fluid and plasmatic marker for the diagnosis of Creutzfeldt-Jakob disease. Proteomics **3**(8): 1495-9.

46. **Gygi, S. P., G. L. Corthals, Y. Zhang, Y. Rochon, and R. Aebersold**. 2000. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc Natl Acad Sci U S A **97**(17): 9390-5.

47. **Hall, D. A., H. Zhu, X. Zhu, T. Royce, M. Gerstein, and M. Snyder**. 2004. Regulation of gene expression by a metabolic enzyme. Science **306**(5695): 482-4.

48. **Hamasur, B., J. Bruchfeld, M. Haile, A. Pawlowski, B. Bjorvatn, G. Kallenius, and S. B. Svenson**. 2001. Rapid diagnosis of tuberculosis by detection of mycobacterial lipoarabinomannan in urine. J Microbiol Methods **45**(1): 41-52.

49. **Hanna, B. A.**, Laboratory Diagnosis, in Tuberculosis, S. Garay and W. Rom, Editors. 2004, Lippincott, Williams, & Wilkins: Philadelphia. p. 163-76.

50. **Haqqani, A. S., J. S. Hutchison, R. Ward, and D. B. Stanimirovic**. 2007. Biomarkers and diagnosis; protein biomarkers in serum of pediatric patients with severe traumatic brain injury identified by ICAT-LC-MS/MS. J Neurotrauma **24**(1): 54-74.

51. **Hellyer, T. J., L. E. DesJardin, L. Teixeira, M. D. Perkins, M. D. Cave, and K. D. Eisenach**. 1999. Detection of viable *Mycobacterium tuberculosis* by

reverse transcriptase-strand displacement amplification of mRNA. J Clin Microbiol **37**(3): 518-23.

52. **Hiller, R., S. Laffer, C. Harwanegg, M. Huber, W. M. Schmidt, A. Twardosz, B. Barletta, W. M. Becker, K. Blaser, H. Breiteneder, M. Chapman, R. Crameri, M. Duchene, F. Ferreira, H. Fiebig, K. Hoffmann-Sommergruber, T. P. King, T. Kleber-Janke, V. P. Kurup, S. B. Lehrer, J. Lidholm, U. Muller, C. Pini, G. Reese, O. Scheiner, A. Scheynius, H. D. Shen, S. Spitzauer, R. Suck, I. Swoboda, W. Thomas, R. Tinghino, M. Van Hage-Hamsten, T. Virtanen, D. Kraft, M. W. Muller, and R. Valenta**. 2002. Microarrayed allergen molecules: diagnostic gatekeepers for allergy treatment. Faseb J **16**(3): 414-6.

53. **Hlavaty, J. J., A. W. Partin, M. J. Shue, L. A. Mangold, J. Derby, T. Javier, S. Kelley, A. Stieg, J. V. Briggman, G. M. Hass, and Y. J. Wu**. 2003. Identification and preliminary clinical evaluation of a 50.8-kDa serum marker for prostate cancer. Urology **61**(6): 1261-5.

54. **Hu, S., J. A. Loo, and D. T. Wong**. 2006. Human body fluid proteome analysis. Proteomics **6**(23): 6326-53.

55. **Huang, J., H. Zhu, S. J. Haggarty, D. R. Spring, H. Hwang, F. Jin, M. Snyder, and S. L. Schreiber**. 2004. Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. Proc Natl Acad Sci U S A **101**(47): 16594-9.

56. **Hueber, W. and W. H. Robinson**. 2006. Proteomic biomarkers for autoimmune disease. Proteomics **6**(14): 4100-5.

57. **Immanuel, C., G. S. Acharyulu, M. Kannapiran, R. Segaran, and G. R. Sarma**. 1990. Acute phase proteins in tuberculous patients. Indian J Chest Dis Allied Sci **32**(1): 15-23.

58. **Issaq, H. J. and T. D. Veenstra**. 2007. The role of electrophoresis in disease biomarker discovery. Electrophoresis **28**(12): 1980-8.

59. **Jindani, A., V. R. Aber, E. A. Edwards, and D. A. Mitchison**. 1980. The early bactericidal activity of drugs in patients with pulmonary tuberculosis. Am Rev Respir Dis **121**(6): 939-49.

60. **Jones, R. B., A. Gordus, J. A. Krall, and G. MacBeath**. 2006. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature **439**(7073): 168-74.

61. **Joos, T. O., M. Schrenk, P. Hopfl, K. Kroger, U. Chowdhury, D. Stoll, D. Schorner, M. Durr, K. Herick, S. Rupp, K. Sohn, and H. Hammerle**. 2000. A

microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. Electrophoresis **21**(13): 2641-50.

62. **Kashyap, R. S., K. M. Dobos, J. T. Belisle, H. J. Purohit, N. H. Chandak, G. M. Taori, and H. F. Daginawala**. 2005. Demonstration of components of antigen 85 complex in cerebrospinal fluid of tuberculous meningitis patients. Clin Diagn Lab Immunol **12**(6): 752-8.

63. **Kashyap, R. S., A. N. Rajan, S. S. Ramteke, V. S. Agrawal, S. S. Kelkar, H. J. Purohit, G. M. Taori, and H. F. Daginawala**. 2007. Diagnosis of tuberculosis in an Indian population by an indirect ELISA protocol based on detection of Antigen 85 complex: a prospective cohort study. BMC Infect Dis **7**: 74.

64. **Kennedy, N., S. H. Gillespie, A. O. Saruni, G. Kisyombe, R. McNerney, F. I. Ngowi, and S. Wilson**. 1994. Polymerase chain reaction for assessing treatment response in patients with pulmonary tuberculosis. J Infect Dis **170**(3): 713-6.

65. **Khan, A. and N. H. Packer**. 2006. Simple urinary sample preparation for proteomic analysis. J Proteome Res **5**(10): 2824-38.

66. **Kuhn, E., J. Wu, J. Karl, H. Liao, W. Zolg, and B. Guild**. 2004. Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and 13C-labeled peptide standards. Proteomics **4**(4): 1175-86.

67. **Laal, S., K. M. Samanich, M. G. Sonnenberg, S. Zolla-Pazner, J. M. Phadtare, and J. T. Belisle**. 1997. Human humoral responses to antigens of *Mycobacterium tuberculosis*: immunodominance of high-molecular-mass antigens. Clin Diagn Lab Immunol **4**(1): 49-56.

68. **Landowski, C. P., H. P. Godfrey, S. I. Bentley-Hibbert, X. Liu, Z. Huang, R. Sepulveda, K. Huygen, M. L. Gennaro, F. H. Moy, S. A. Lesley, and M. Haak-Frendscho**. 2001. Combinatorial use of antibodies to secreted mycobacterial proteins in a host immune system-independent test for tuberculosis. J Clin Microbiol **39**(7): 2418-24.

69. **Le Naour, F.** 2007. Identification of tumor antigens by using proteomics. Methods Mol Biol **360**: 327-34.

70. **Lehrer, S., J. Roboz, H. Ding, S. Zhao, E. J. Diamond, J. F. Holland, N. N. Stone, M. J. Droller, and R. G. Stock**. 2003. Putative protein markers in the sera of men with prostatic neoplasms. BJU Int **92**(3): 223-5.

71. **Lescuyer, P., D. Hochstrasser, and T. Rabilloud**. 2007. How Shall We Use the Proteomics Toolbox for Biomarker Discovery? J Proteome Res **6**(9): 3371-6.

72. **Levin, Y., E. Schwarz, L. Wang, F. M. Leweke, and S. Bahn**. 2007. Label-free LC-MS/MS quantitative proteomics for large-scale biomarker discovery in complex samples. J Sep Sci **30**(14): 2198-203.

73. **Li, B., L. Jiang, Q. Song, J. Yang, Z. Chen, Z. Guo, D. Zhou, Z. Du, Y. Song, J. Wang, H. Wang, S. Yu, and R. Yang**. 2005. Protein microarray for profiling antibody responses to Yersinia pestis live vaccine. Infect Immun **73**(6): 3734-9.

74. **Li, Q. Z., C. Xie, T. Wu, M. Mackay, C. Aranow, C. Putterman, and C. Mohan**. 2005. Identification of autoantibody clusters that best predict lupus disease activity using glomerular proteome arrays. J Clin Invest **115**(12): 3428-39.

75. **Li, X. J., E. C. Yi, C. J. Kemp, H. Zhang, and R. Aebersold**. 2005. A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography-mass spectrometry. Mol Cell Proteomics **4**(9): 1328-40.

76. **Liu, H. C., C. J. Hu, J. G. Chang, S. M. Sung, L. S. Lee, R. Y. Yuan, and S. J. Leu**. 2006. Proteomic identification of lower apolipoprotein A-I in Alzheimer's disease. Dement Geriatr Cogn Disord **21**(3): 155-61.

77. **Lueking, A., O. Huber, C. Wirths, K. Schulte, K. M. Stieler, U. Blume-Peytavi, A. Kowald, K. Hensel-Wiegel, R. Tauber, H. Lehrach, H. E. Meyer, and D. J. Cahill**. 2005. Profiling of alopecia areata autoantigens based on protein microarray technology. Mol Cell Proteomics **4**(9): 1382-90.

78. **Lueking, A., A. Possling, O. Huber, A. Beveridge, M. Horn, H. Eickhoff, J. Schuchardt, H. Lehrach, and D. J. Cahill**. 2003. A nonredundant human protein chip for antibody screening and serum profiling. Mol Cell Proteomics **2**(12): 1342-9.

79. **Makretsov, N. A., D. G. Huntsman, T. O. Nielsen, E. Yorida, M. Peacock, M. C. Cheang, S. E. Dunn, M. Hayes, M. van de Rijn, C. Bajdik, and C. B. Gilks**. 2004. Hierarchical clustering analysis of tissue microarray immunostaining data identifies prognostically significant groups of breast carcinoma. Clin Cancer Res **10**(18 Pt 1): 6143-51.

80. **Marshall, J., P. Kupchak, W. Zhu, J. Yantha, T. Vrees, S. Furesz, K. Jacks, C. Smith, I. Kireeva, R. Zhang, M. Takahashi, E. Stanton, and G. Jackowski**. 2003. Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. J Proteome Res **2**(4): 361-72.

81. **Mehta, A. I., S. Ross, M. S. Lowenthal, V. Fusaro, D. A. Fishman, E. F. Petricoin, 3rd, and L. A. Liotta**. 2003. Biomarker amplification by serum carrier protein binding. Dis Markers **19**(1): 1-10.

82. **Michaud, G. A., M. Salcius, F. Zhou, R. Bangham, J. Bonin, H. Guo, M. Snyder, P. F. Predki, and B. I. Schweitzer**. 2003. Analyzing antibody specificity with whole proteome microarrays. Nat Biotechnol **21**(12): 1509-12.

83. **Miller, J. C., H. Zhou, J. Kwekel, R. Cavallo, J. Burke, E. B. Butler, B. S. Teh, and B. B. Haab**. 2003. Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. Proteomics **3**(1): 56-63.

84. **Mitchison, D. A.** 1993. Assessment of new sterilizing drugs for treating pulmonary tuberculosis by culture at 2 months. Am Rev Respir Dis **147**(4): 1062-3.

85. **Moore, D. F., J. I. Curry, C. A. Knott, and V. Jonas**. 1996. Amplification of rRNA for assessment of treatment response of pulmonary tuberculosis patients during antimicrobial therapy. J Clin Microbiol **34**(7): 1745-9.

86. **Neuman de Vegvar, H. E., R. R. Amara, L. Steinman, P. J. Utz, H. L. Robinson, and W. H. Robinson**. 2003. Microarray profiling of antibody responses against simian-human immunodeficiency virus: postchallenge convergence of reactivities independent of host histocompatibility type and vaccine regimen. J Virol **77**(20): 11125-38.

87. **Omenn, G. S., D. J. States, M. Adamski, T. W. Blackwell, R. Menon, H. Hermjakob, R. Apweiler, B. B. Haab, R. J. Simpson, J. S. Eddes, E. A. Kapp, R. L. Moritz, D. W. Chan, A. J. Rai, A. Admon, R. Aebersold, J. Eng, W. S. Hancock, S. A. Hefta, H. Meyer, Y. K. Paik, J. S. Yoo, P. Ping, J. Pounds, J. Adkins, X. Qian, R. Wang, V. Wasinger, C. Y. Wu, X. Zhao, R. Zeng, A. Archakov, A. Tsugita, I. Beer, A. Pandey, M. Pisano, P. Andrews, H. Tammen, D. W. Speicher, and S. M. Hanash**. 2005. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics **5**(13): 3226-45.

88. **Orchekowski, R., D. Hamelinck, L. Li, E. Gliwa, M. vanBrocklin, J. A. Marrero, G. F. Vande Woude, Z. Feng, R. Brand, and B. B. Haab**. 2005. Antibody microarray profiling reveals individual and combined serum proteins associated with pancreatic cancer. Cancer Res **65**(23): 11193-202.

89. **Pawlik, T. M., D. H. Hawke, Y. Liu, S. Krishnamurthy, H. Fritsche, K. K. Hunt, and H. M. Kuerer**. 2006. Proteomic analysis of nipple aspirate fluid from women with early-stage breast cancer using isotope-coded affinity tags and tandem mass spectrometry reveals differential expression of vitamin D binding protein. BMC Cancer **6**: 68.

90. **Pereira Arias-Bouda, L. M., L. N. Nguyen, L. M. Ho, S. Kuijper, H. M. Jansen, and A. H. Kolk**. 2000. Development of antigen detection assay for diagnosis of tuberculosis using sputum samples. J Clin Microbiol **38**(6): 2278-83.

91. **Perrin, F. M., M. C. Lipman, T. D. McHugh, and S. H. Gillespie**. 2007. Biomarkers of treatment response in clinical trials of novel antituberculosis agents. Lancet Infect Dis **7**(7): 481-90.

92. **Petricoin, E., 3rd and L. A. Liotta**. 2003. Counterpoint: The vision for a new diagnostic paradigm. Clin Chem **49**(8): 1276-8.

93. **Petricoin, E. F., A. M. Ardekani, B. A. Hitt, P. J. Levine, V. A. Fusaro, S. M. Steinberg, G. B. Mills, C. Simone, D. A. Fishman, E. C. Kohn, and L. A. Liotta**. 2002. Use of proteomic patterns in serum to identify ovarian cancer. Lancet **359**(9306): 572-7.

94. **Petricoin, E. F., C. Belluco, R. P. Araujo, and L. A. Liotta**. 2006. The blood peptidome: a higher dimension of information content for cancer biomarker discovery. Nat Rev Cancer **6**(12): 961-7.

95. **Pieper, R., C. L. Gatlin, A. M. McGrath, A. J. Makusky, M. Mondal, M. Seonarain, E. Field, C. R. Schatz, M. A. Estock, N. Ahmed, N. G. Anderson, and S. Steiner**. 2004. Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. Proteomics **4**(4): 1159-74.

96. **Pisitkun, T., R. Johnstone, and M. A. Knepper**. 2006. Discovery of urinary biomarkers. Mol Cell Proteomics **5**(10): 1760-71.

97. **Pitarch, A., J. Abian, M. Carrascal, M. Sanchez, C. Nombela, and C. Gil**. 2004. Proteomics-based identification of novel Candida albicans antigens for diagnosis of systemic candidiasis in patients with underlying hematological malignancies. Proteomics **4**(10): 3084-106.

98. **Powell, K.** 2003. Proteomics delivers on promise of cancer biomarkers. Nat Med **9**(8): 980.

99. **Qiu, J., J. Madoz-Gurpide, D. E. Misek, R. Kuick, D. E. Brenner, G. Michailidis, B. B. Haab, G. S. Omenn, and S. Hanash**. 2004. Development of natural protein microarrays for diagnosing cancer based on an antibody response to tumor antigens. J Proteome Res **3**(2): 261-7.

100. **Ribeiro-Rodrigues, R., T. Resende Co, J. L. Johnson, F. Ribeiro, M. Palaci, R. T. Sa, E. L. Maciel, F. E. Pereira Lima, V. Dettoni, Z. Toossi, W. H. Boom, R. Dietze, J. J. Ellner, and C. S. Hirsch**. 2002. Sputum cytokine levels in

patients with pulmonary tuberculosis as early markers of mycobacterial clearance. Clin Diagn Lab Immunol **9**(4): 818-23.

101. **Rifai, N., M. A. Gillette, and S. A. Carr**. 2006. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol **24**(8): 971-83.

102. **Roxo-Rosa, M., G. da Costa, T. M. Luider, B. J. Scholte, A. V. Coelho, M. D. Amaral, and D. Penque**. 2006. Proteomic analysis of nasal cells from cystic fibrosis patients and non-cystic fibrosis control individuals: search for novel biomarkers of cystic fibrosis lung disease. Proteomics **6**(7): 2314-25.

103. **Samanich, K., J. T. Belisle, and S. Laal**. 2001. Homogeneity of antibody responses in tuberculosis patients. Infect Immun **69**(7): 4600-9.

104. **Samanich, K. M., J. T. Belisle, M. G. Sonnenberg, M. A. Keen, S. Zolla-Pazner, and S. Laal**. 1998. Delineation of human antibody responses to culture filtrate antigens of *Mycobacterium tuberculosis*. J Infect Dis **178**(5): 1534-8.

105. **Schena, M., D. Shalon, R. W. Davis, and P. O. Brown**. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270**(5235): 467-70.

106. **Schluger, N. W.**, Novel Approaches to the Rapid Diagnosis of Tuberculosis, in Tuberculosis, S. Garay and W. Rom, Editors. 2004, Lippincott, Williams, & Wilkins: Philadelphia. p. 177-81.

107. **Simpkins, F., J. A. Czechowicz, L. Liotta, and E. C. Kohn**. 2005. SELDI-TOF mass spectrometry for cancer biomarker discovery and serum proteomic diagnostics. Pharmacogenomics **6**(6): 647-53.

108. **Sinz, A., M. Bantscheff, S. Mikkat, B. Ringel, S. Drynda, J. Kekow, H. J. Thiesen, and M. O. Glocker**. 2002. Mass spectrometric proteome analyses of synovial fluids and plasmas from patients suffering from rheumatoid arthritis and comparison to reactive arthritis or osteoarthritis. Electrophoresis **23**(19): 3445-56.

109. **Soen, Y., D. S. Chen, D. L. Kraft, M. M. Davis, and P. O. Brown**. 2003. Detection and characterization of cellular immune responses using peptide-MHC microarrays. PLoS Biol **1**(3): E65.

110. **Speer, R., J. D. Wulfkuhle, L. A. Liotta, and E. F. Petricoin, 3rd**. 2005. Reverse-phase protein microarrays for tissue-based analysis. Curr Opin Mol Ther 7(3): 240-5.

111. **Steller, S., P. Angenendt, D. J. Cahill, S. Heuberger, H. Lehrach, and J. Kreutzberger**. 2005. Bacterial protein microarrays for identification of new potential diagnostic markers for Neisseria meningitidis infections. Proteomics **5**(8): 2048-55.

112. **Stone, J. D., W. E. Demkowicz, Jr., and L. J. Stern**. 2005. HLA-restricted epitope identification and detection of functional T cell responses by using MHC-peptide and costimulatory microarrays. Proc Natl Acad Sci U S A **102**(10): 3744-9.

113. **Sundaresh, S., A. Randall, B. Unal, J. M. Petersen, J. T. Belisle, M. G. Hartley, M. Duffield, R. W. Titball, D. H. Davies, P. L. Felgner, and P. Baldi**. 2007. From protein microarrays to diagnostic antigen discovery: a study of the pathogen Francisella tularensis. Bioinformatics **23**(13): i508-18.

114. **Templin, M. F., D. Stoll, M. Schrenk, P. C. Traub, C. F. Vohringer, and T. O. Joos**. 2002. Protein microarray technology. Trends Biotechnol **20**(4): 160-6.

115. **Tessema, T. A., B. Hamasur, G. Bjun, S. Svenson, and B. Bjorvatn**. 2001. Diagnostic evaluation of urinary lipoarabinomannan at an Ethiopian tuberculosis centre. Scand J Infect Dis **33**(4): 279-84.

116. **Tilleman, K., K. Van Beneden, A. Dhondt, I. Hoffman, F. De Keyser, E. Veys, D. Elewaut, and D. Deforce**. 2005. Chronically inflamed synovium from spondyloarthropathy and rheumatoid arthritis investigated by protein expression profiling followed by tandem mass spectrometry. Proteomics **5**(8): 2247-57.

117. **Veenstra, T. D., T. P. Conrads, B. L. Hood, A. M. Avellino, R. G. Ellenbogen, and R. S. Morrison**. 2005. Biomarkers: mining the biofluid proteome. Mol Cell Proteomics **4**(4): 409-18.

118. **Villanueva, J., D. R. Shaffer, J. Philip, C. A. Chaparro, H. Erdjument-Bromage, A. B. Olshen, M. Fleisher, H. Lilja, E. Brogi, J. Boyd, M. Sanchez-Carbayo, E. C. Holland, C. Cordon-Cardo, H. I. Scher, and P. Tempst**. 2006. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest **116**(1): 271-84.

119. **Villar-Garea, A., M. Griese, and A. Imhof**. 2007. Biomarker discovery from body fluids using mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci **849**(1-2): 105-14.

120. **Vitzthum, F., F. Behrens, N. L. Anderson, and J. H. Shaw**. 2005. Proteomics: from basic research to diagnostic application. A review of requirements & needs. J Proteome Res **4**(4): 1086-97.

121. **Wallis, R. S., M. Perkins, M. Phillips, M. Joloba, B. Demchuk, A. Namale, J. L. Johnson, D. Williams, K. Wolski, L. Teixeira, R. Dietze, R. D. Mugerwa, K. Eisenach, and J. J. Ellner**. 1998. Induction of the antigen 85 complex of *Mycobacterium tuberculosis* in sputum: a determinant of outcome in pulmonary tuberculosis treatment. J Infect Dis **178**(4): 1115-21.

122. **Wallis, R. S., S. A. Vinhas, J. L. Johnson, F. C. Ribeiro, M. Palaci, R. L. Peres, R. T. Sa, R. Dietze, A. Chiunda, K. Eisenach, and J. J. Ellner**. 2003. Whole blood bactericidal activity during treatment of pulmonary tuberculosis. J Infect Dis **187**(2): 270-8.

123. **Washburn, M. P., D. Wolters, and J. R. Yates, 3rd**. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol **19**(3): 242-7.

124. **Weldingh, K., I. Rosenkrands, S. Jacobsen, P. B. Rasmussen, M. J. Elhay, and P. Andersen**. 1998. Two-dimensional electrophoresis for analysis of *Mycobacterium tuberculosis* culture filtrate and purification and characterization of six novel proteins. Infect Immun **66**(8): 3492-500.

125. **Wu, L. and D. K. Han**. 2006. Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics. Expert Rev Proteomics **3**(6): 611-9.

126. **Xiao, Z., B. L. Adam, L. H. Cazares, M. A. Clements, J. W. Davis, P. F. Schellhammer, E. A. Dalmasso, and G. L. Wright, Jr.** 2001. Quantitation of serum prostate-specific membrane antigen by a novel protein biochip immunoassay discriminates benign from malignant prostate disease. Cancer Res **61**(16): 6029-33.

127. **Yates, J. R., 3rd**. 2004. Mass spectral analysis in proteomics. Annu Rev Biophys Biomol Struct **33**: 297-316.

128. **Ye, B., D. W. Cramer, S. J. Skates, S. P. Gygi, V. Pratomo, L. Fu, N. K. Horick, L. J. Licklider, J. O. Schorge, R. S. Berkowitz, and S. C. Mok**. 2003. Haptoglobin-alpha subunit as potential serum biomarker in ovarian cancer: identification and characterization using proteomic profiling and mass spectrometry. Clin Cancer Res **9**(8): 2904-11.

129. **Zhang, H., E. C. Yi, X. J. Li, P. Mallick, K. S. Kelly-Spratt, C. D. Masselon, D. G. Camp, 2nd, R. D. Smith, C. J. Kemp, and R. Aebersold**. 2005. High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. Mol Cell Proteomics **4**(2): 144-55.

130. **Zhang, Z., R. C. Bast, Jr., Y. Yu, J. Li, L. J. Sokoll, A. J. Rai, J. M. Rosenzweig, B. Cameron, Y. Y. Wang, X. Y. Meng, A. Berchuck, C. Van Haaften-Day, N. F. Hacker, H. W. de Bruijn, A. G. van der Zee, I. J. Jacobs,**

E. T. Fung, and D. W. Chan. 2004. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Res **64**(16): 5882-90.

131. **Zhou, M., D. A. Lucas, K. C. Chan, H. J. Issaq, E. F. Petricoin, 3rd, L. A. Liotta, T. D. Veenstra, and T. P. Conrads**. 2004. An investigation into the human serum "interactome". Electrophoresis **25**(9): 1289-98.

132. **Zhu, H., M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein, and M. Snyder**. 2001. Global analysis of protein activities using proteome chips. Science **293**(5537): 2101-5.

133. **Zhu, H., S. Hu, G. Jona, X. Zhu, N. Kreiswirth, B. M. Willey, T. Mazzulli, G. Liu, Q. Song, P. Chen, M. Cameron, A. Tyler, J. Wang, J. Wen, W. Chen, S. Compton, and M. Snyder**. 2006. Severe acute respiratory syndrome diagnostics using a coronavirus protein microarray. Proc Natl Acad Sci U S A **103**(11): 4011-6.

134. **Zhu, H., J. F. Klemic, S. Chang, P. Bertone, A. Casamayor, K. G. Klemic, D. Smith, M. Gerstein, M. A. Reed, and M. Snyder**. 2000. Analysis of yeast protein kinases using protein chips. Nat Genet **26**(3): 283-9.

135. **Zolg, J. W. and H. Langen**. 2004. How industry is approaching the search for new diagnostic markers and biomarkers. Mol Cell Proteomics **3**(4): 345-54.

136. **Zolg, W.** 2006. The proteomic search for diagnostic biomarkers: lost in translation? Mol Cell Proteomics **5**(10): 1720-6.

Chapter III

Disease State Differentiation and Evaluation of Tuberculosis Biomarkers via Native Antigen Array Profiling

Partially presented in Mark J. Sartain, Richard A. Slayden, Krishna K. Singh, Suman Laal, and John T. Belisle. 2006. Disease State Differentiation and Identification of Tuberculosis Biomarker via Native Antigen Array Profiling. Molecular and Cellular Proteomics. Nov; Vol 5 (11): 2102-2113

3.1 Introduction

In recent years there has been renewed interest in developing antibody-based diagnostics that utilize multiple antigens to achieve high levels of sensitivity and specificity (7). The success of a serodiagnostic test for TB hinges on its ability to detect multiple disease states, including pauci- and multibacillary forms, pediatric cases, and patients coinfected with HIV. Previous work from our laboratories identified several antigens that provide high sensitivity and specificity when used in an ELISA format (5, 6, 16-20). Furthermore, this work highlighted differential antigen reactivity based on the disease state (5, 15-18, 20). However, a complete analysis of patients' serological reactivity to a large proportion of the *Mtb* proteome is hindered by the inability to evaluate the reactivity of the nearly 4,000 predicted proteins of *Mtb* in a high-throughput fashion.

In the absence of a complete *Mtb* ORF library, methods to produce a first generation *Mtb* protein microarray based on native proteins were required. A multi-dimensional separation strategy was devised to efficiently resolve native proteins found in the cytosol and CF of *Mtb.* This resulted in 960 relatively simple protein fractions from two highly complex protein pools. These fractions were spotted to nitrocellulose slides and probed with sera from PPD-positive (PPD+) healthy controls, cavitary TB,

noncavitary TB, HIV+TB+ patients. Overall the analyses 1) revealed distinct trends in patient reactivity profiles, 2) corroborated our initial findings from earlier 2-D immunoblot based experiments, and 3) demonstrated the raw power of this novel methodology in the high-throughput characterization of TB antigens. Furthermore, four proteins specific for cavitary TB patients were identified, and four novel antigens previously undetected by other methods were defined as serodiagnostic targets.

3.2 Materials and methods

3.2.1 Preparation of *Mtb* subcellular fractions

Mtb strain H37Rv was expanded from a 1-ml frozen stock (approximately 10^8 colony-forming units per ml) to 24 L of late log culture in glycerol-alanine-salts (GAS) medium (22). The culture supernatant was separated from the cells and processed to generate the CFP of *Mtb* as previously described (2). The *Mtb* H37Rv cells (88.9 g wet weight) were washed 3 times with PBS (pH 7.4), frozen at -70°C and inactivated with 24 kilograys of γ-irradiation. Lysis of these cells was achieved by suspending in 44 ml of TSE buffer [10 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA] containing 0.06% DNase, 0.06% RNase, 0.07% pepstatin, 0.05% leupeptin and 20 μM PMSF and passing through a French Press five times at 1500 psi. The resulting lysate was diluted with 1 vol of TSE buffer and centrifuged at 2,000 x *g* for 5 min to remove unbroken cells. The cytosol was obtained as the final supernatant of sequential centrifugations at 27,000 x g and 100,000 x *g* (4), and was dialyzed against 10 mM ammonium bicarbonate using a 3,500 Da molecular mass cut-off membrane. The protein concentrations of the cytosol and CFP were determined with the bicinchoninic acid (BCA) protein assay (21).

3.2.2 Multi-dimensional protein fractionation

Initial fractionation of the CFP (124 ml at 3.6 mg/ml) and cytosolic proteins (200 ml at 2.5 mg/ml) was achieved with sequential rounds of ammonium sulfate precipitation. Specifically, the CFP and cytosolic proteins were precipitated with 42% and 67%; and 29% and 44% saturated ammonium sulfate, respectively. Precipitated proteins were collected by centrifugation at 10,000 x *g*, 4°C for 1 hr. All protein pellets were suspended in 20 mM Tris-HCl (pH 8.0). These suspensions and the final supernatants of the sequential precipitations were dialyzed against 20 mM Tris-HCl (pH 8.0), using a 3,500 Da molecular weight cutoff membrane, and the dialyzed protein solutions were concentrated where needed. To ensure the removal of contaminating nucleic acids, $MgCl_2$ (5 mM final conc.) and DNase and RNase (1.25% final conc.) were added to each fraction followed by incubation at 37°C for 30 min.

The fractions obtained by ammonium sulfate precipitation were adjusted to 10% acetonitrile, and applied to an HPLC column (1 x 10 cm) packed with Source 15Q strong anion-exchange (AIEX) resin (Amersham Biosciences, Piscataway, NJ). Proteins were eluted with a step gradient of increasing concentrations of NaCl at a flow rate of 3.3 ml per min using a Waters 600E HPLC system (Waters Corp., Milford, MA). The eluted protein fractions were concentrated 100-fold, and exchanged into 20 mM ammonium bicarbonate by ultrafiltration. Protein concentrations were determined by the BCA protein assay, and fractions containing less than 1 mg protein were pooled. All other fractions were kept separate. The concentrated AIEX fractions were adjusted to 10% acetonitrile, applied to a HPLC Source™ 15RPC ST 4.6/100 column (Amersham) and the proteins eluted with an increasing linear gradient (10% to 70%) of acetonitrile. All

fractions were dried in a speed-vac, suspended in 67 μl of 10 mM ammonium bicarbonate, and protein concentrations determined by the BCA protein assay.

3.2.3 Human sera and antibodies

Sera from the following groups of individuals were obtained by our collaborator Dr. Suman Laal (NYU) with informed consent.

(i) *Twelve PPD+ healthy individuals.* Seven of these individuals were recent immigrants from countries where *Mtb* is endemic, many of whom had been vaccinated with *M. bovis* BCG. The remaining five PPD+ healthy individuals were from the United States or western Europe and were not BCG-vaccinated.

(ii) *Nine noncavitary TB patients with no recognizable cavitary lesions on chest X rays.* These were AFB sputum-smear-negative (6/9) or positive (3/9), culture positive patients attending the infectious disease clinic at the Manhattan VA Medical Center (MVAMC). None of these patients were HIV-infected. These individuals were bled either prior to or within two weeks of the initiation of therapy for TB.

(iii) *Eleven cavitary TB patients, with moderate-to-advanced cavitary lesions as determined by chest X rays.* These were AFB sputum-smear positive patients obtained from the Lala Ram Sarup Institute of Tuberculosis and Respiratory Diseases (LRSITRD) in New Delhi, India who were all bled prior to initiation of therapy for TB. None of these patients were HIV infected.

(iv) *Ten HIV+TB+ patients.* These were sputum smear positive (7/10) or negative (3/10), culture-confirmed patients from the MVAMC. None of the patients had radiological evidence of cavitary lesions. All ten patients were known to possess

antibodies to the CFP of *Mtb* when tested by ELISA in earlier studies (5, 17). These patients were bled either prior to or within two weeks of the initiation of therapy for TB.

(v) *Six HIV+TB- patients.* These were asymptomatic, HIV-infected individuals from the MVAMC. All sera were preadsorbed with *E. coli* lysates to remove cross-reactive antibodies to ubiquitous prokaryotic proteins as described earlier (6).

Monoclonal antibodies (MAbs) and polyclonal sera against specific *Mtb* proteins were obtained from the Colorado State University TB Research Materials and Vaccine Testing Contract (NIH, NIAID NO1-AI-75320). The following antibodies and dilutions were used for both microarray analyses and immunoblots: IT-12 α-19 kDa (1:20), IT-20 α-HspX (1:100), IT-23 α-PstS1 (1:20), IT-47 α-PstS1 (1:20), IT-52 α-MPT51 (1:5), CS-35 α-LAM (1:20), CS-49 α-HspX (1:100), CS-93 α-45 kDa (1:20), and α-45 kDa polyclonal sera (1:1000).

3.2.4 Protein microarray printing and probing

An aliquot (5 μg protein) of each multi-dimensional chromatography fraction was transferred to 384-well microtiter plates, dried, and solubilized in 25 μl FAST® protein array print buffer (Schleicher & Schuell Bioscience, Keene, NH). The plates were centrifuged briefly (2,000 x g) to pellet any precipitate, and ~1 nl of the supernatants (0.2 mg per ml) was printed to nitrocellulose-coated FAST® glass slides (Schleicher & Schuell) using Stealth SMP3® spotting pins (TeleChem International, Sunnyvale, CA, www.arrayit.com) and a VersaArray® Chipwriter Pro microarray contact printer (Bio-Rad Laboratories, Hercules, CA). Cytosolic proteins, CFP, the native 38-kDa PstS1 protein (Rv0934), and the six ammonium sulfate precipitation fractions were also printed in a dilution series of 1.6, 0.8, 0.4, 0.2, 0.1, 0.5, 0.25, and 0.125 mg/ml. As negative

controls, *E. coli* WCL was printed in the same dilution series, and FAST® print buffer was printed alone. All samples were printed in triplicate, resulting in 3,768 total spots per slide. The slides were allowed to dry 1 hour at room temperature (RT) and stored at 4°C until use. Printed microarray slides were washed 10 min in commercial FAST® protein array wash buffer (Schleicher & Schuell), and probed with individual serum (750 µl) diluted 1:100 in PBS (pH 7.4), 1% BSA for 1 h at RT. Slides were washed twice for 10 min in FAST® protein array wash buffer and probed for 1 h at RT with Cy3-conjugated anti-human IgG (Sigma, St. Louis, MO) diluted 1:500 in FAST® protein array wash buffer. Slides were again washed twice for 10 min, allowed to dry, and scanned using a VersArray® Chipreader (Bio-Rad Laboratories, Hercules, CA). Probing of the microarray slides with MAbs or polyclonal sera was performed in the same manner, except Cy3 conjugated anti-mouse IgG or Cy5 conjugated anti-rabbit IgG (Amersham Biosciences) were used, respectively, as the secondary antibody. Microarray analyses of individual patient's serum were repeated in triplicate, and one slide was used for each MAb or rabbit polyclonal sera.

3.2.5 Microarray data analyses

Microarray spot intensity values were quantified with TIGR Spotfinder software (14). Signal-to-Noise Ratios (SNRs) were calculated for each spot by dividing the raw intensity (sum of all pixels per spot) by the background intensity (local background median multiplied by spot area). Analysis of SNR reduced the local background variation or bias observed between individual slides. The mean SNR for each protein or protein fraction printed in triplicate was determined, resulting in an averaged SNR (AvSNR). To allow for direct patient-to-patient or slide-to-slide comparisons, all

AvSNRs for a slide were normalized against the median AvSNR of all multidimensional chromatography fractions of the slide. The normalized AvSNR (NAvSNR) was based on the median AvSNR rather than the mean AvSNR since the median AvSNR was less affected by variations in reactivity between sera. For the microarray slides probed with MAbs or rabbit polyclonal sera, the AvSNR for each fraction was calculated.

3.2.6 SDS-PAGE and Western blot analyses

SDS-PAGE of multidimensional chromatography fractions was performed with 10% to 20% polyacrylamide gradient Tricine gels (Invitrogen, Carlsbad, CA) or 15% Tris-glycine gels (10 cm x 7.5 cm) (8). Protein staining was achieved with silver nitrate (12) or Coomassie Brilliant Blue R-250. For Western blot analyses, aliquots (3 μg) of selected fractions were resolved on 10% to 20% polyacrylamide gradient Tricine gels and electroblotted to nitrocellulose membranes (22). The membranes were blocked with 3% nonfat milk in PBS (pH 7.2) for 2 h, washed with PBS containing 2% Tween 20, and exposed overnight to preabsorbed pooled sera from patients and control subjects diluted 1:200. The blots were washed with PBS containing 2% Tween 20, probed with alkaline phosphatase-conjugated anti-human IgG (1:2,000, Sigma) for 1.5 h, and washed extensively. Antigen-antibody complexes were visualized by color development with 5-bromo-4-chloro-indoyl-phosphatase-nitroblue tetrazolium substrate (Kirkegaard & Perry Laboratories, Gaithersburg, MD).

3.2.7 Mass spectrometry

Coomassie stained protein bands corresponding to those that reacted to patients' sera on Western blots were excised, subjected to in-gel digestion with modified trypsin, chymotrypsin, or endoproteinase GluC (Roche Applied Science, Indianapolis, IN), and

the resulting peptides were extracted with 60% acetonitrile, 0.1% TFA (3). Extracted peptides were applied to a capillary (0.2 x 50 mm) C18 reversed phase (RP) column (Microchom BioResources, Auburn, CA) and eluted with an increasing linear gradient (5% to 70%) of acetonitrile in 0.1% acetic acid using an Eldex MicroPro capillary HPLC system (Napa, CA) with a flow rate of 5 μl per min. The RP eluant was introduced directly into a ThermoFinnigan LCQ electrospray mass spectrometer (San Jose, CA) operated using Xcalibur software version 1.3, and the peptides were analyzed by MS/MS. The electrospray needle was set at 4 kV with a N_2 sheath gas flow of 40, and a capillary temperature of 200°C. MS/MS was automatically performed on the most dominant ion of the previous scan and the normalized collision energy was set at 40%. BioWorks 3.1 turboSEQUEST software (ThermoFinnigan) was used to match the MS/MS data of peptides to protein sequences extracted form the *Mtb* genome database (NC_000962) that contained 3989 proteins. The software was set to evaluate peptides obtained by trypsin or chymotrypsin and GluC digestion and to consider the oxidation of methionine (+16.0 amu) and the acrylamide modification of Cys (+71.0 amu). The Scaffold software (Proteome Software, Portland, OR) that verifies peptide identifications made by SEQUEST and probabilistically validates the peptide and protein identifications was applied to all MS/MS sequencing results.

3.3 Results

3.3.1 Development of a native *Mtb* protein microarray

Recent advances in proteomics have resulted in the ability to separate and identify individual proteins or peptides from complex biological samples. Specifically, multi-dimensional chromatography of peptides, derived from tryptic digests of crude biological

samples, followed by MS/MS analysis (MudPIT) has resulted in the experimental validation of substantial portions of theoretical proteomes (11). Recently, this approach was applied to *Mtb* and resulted in more than a threefold increase in the number of proteins previously identified by 2-D-SDS-PAGE methods (10). With goals similar to the MudPIT strategy, this study investigated methods to efficiently separate complex pools of intact mycobacterial proteins into relatively simple and enriched fractions that could be evaluated for serological reactivity in a high-throughput fashion.

The high-throughput capacity inherent to microarray technology allows for the simultaneous evaluation of upwards to tens of thousands of protein samples. Thus, the limiting factor in reaching the full potential of this technology is the quality, diversity, and sheer number of spotted protein samples. With this consideration in mind, a significant effort was spent developing protein separation strategies with the primary goals being: 1) to produce a large and comprehensive protein set, and 2) to produce a protein set with minimal redundant components. Numerous bioseparation techniques were initially investigated including isoelectric focusing, chromatofocusing, preparative SDS-PAGE w/ whole gel elution, differential ammonium sulfate precipitation, AIEX chromatography, RP chromatography, and hydrophobic interaction chromatography. In addition multiple HPLC column resins were compared and buffer components, buffer pH, and elution gradients were optimized. Protein solubility, protein recovery, and protein separation characteristics were the primary criteria used to select an optimal separation strategy. In the end, a three-dimensional separation scheme was developed. Specifically, aliquots of cytosolic proteins (500 mg) and CFPs (446 mg) from *Mtb* H37Rv were subjected to sequential ammonium sulfate precipitations, resulting in six

fractions with nearly equal protein amounts (Fig. 3.1A). Subsequent AIEX chromatography expanded the fraction number to 78, with protein yields varying from 160 μg to 13 mg per fraction (Fig. 3.1B). Those fractions containing <500 μg of protein were excluded from further separation (8 fractions), and fractions containing between 500 μg and 1 mg were pooled with neighboring fractions (10 fractions). This resulted in a total of 64 fractions that were applied to RP chromatography under mildly basic pH conditions. This multidimensional chromatography separation yielded a total of 960 fractions, and SDS-PAGE analysis revealed varied complexity among the fractions, with 1 to >20 proteins observed per fraction (Fig. 3.1C).

In order to generate a first-generation protein microarray, numerous printing and probing techniques were evaluated. Glass slide surface chemistries (aldehyde, epoxy, amino-silanated, amino-silylated) that bind protein in a covalent manner were evaluated, as well as acrylamide and nitrocellulose substrates that bind protein via hydrophobic interactions. A nitrocellulose-based chemistry was chosen due to the increased likelihood of preserving protein both three-dimensional structures and conformational epitopes, and because conventional immunoblots most commonly utilize nitrocellulose-based supports. Protein printing concentration, spot size, spot spacing, buffer compositions, and antibody concentrations were also optimized. Initial analyses with crude fractions indicated linear relationships for printing concentrations between 0.125 mg/ml and 1.0 mg/ml, and a concentration of 0.2 mg/ml was chosen in an effort to conserve protein (Fig. 3.2A and B).

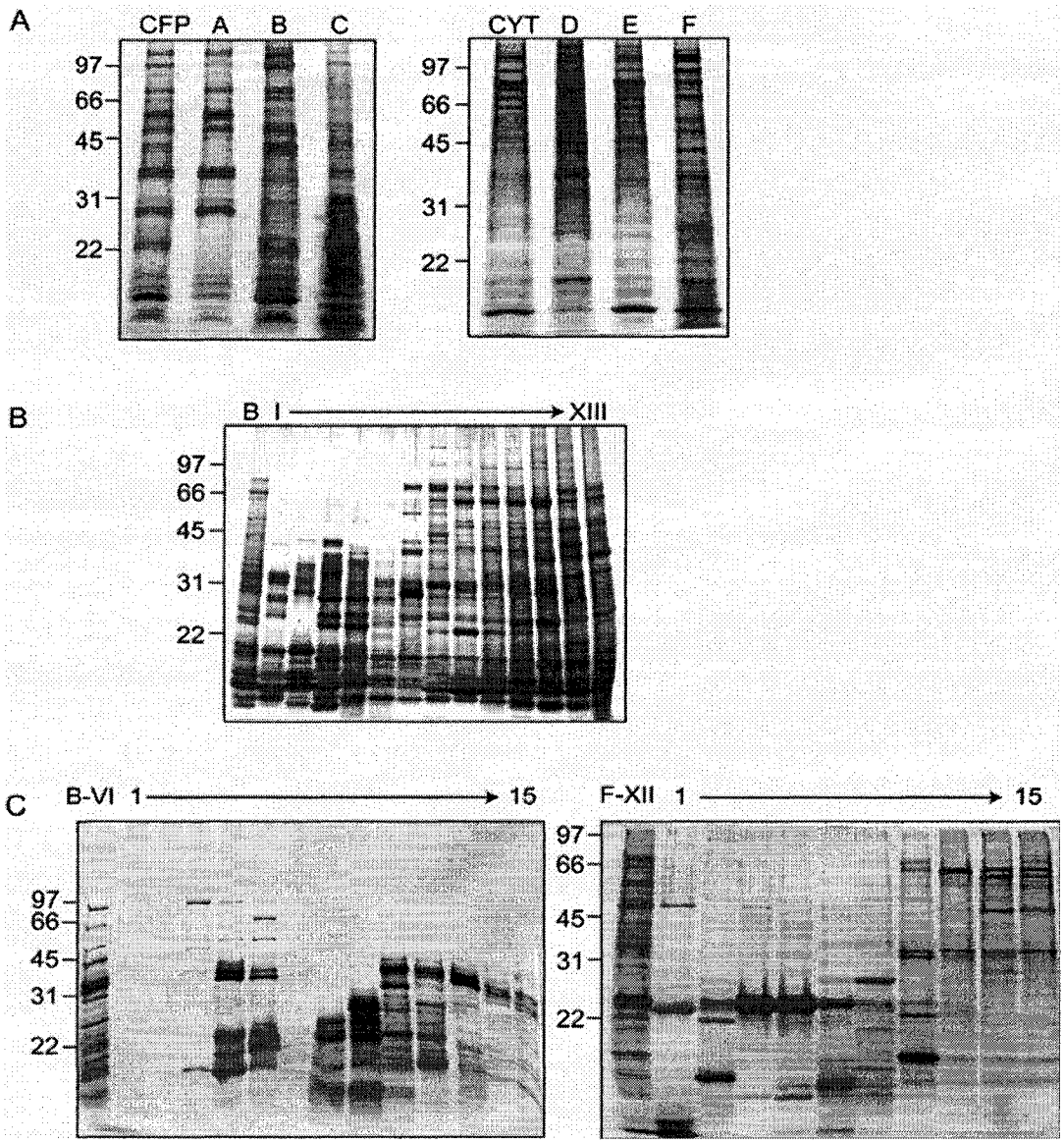


Fig. 3.1 ***Mtb*** **CFP and cytosol proteins were subjected to a multi-dimensional separation scheme.** (**A**) Silver-stained SDS-PAGE analysis of fractionation obtained by ammonium-sulfate precipitation of CFP and cytosol. *CFP*, unfractionated CFP; *lane A*, 42% ammonium sulfate precipitate of the CFP; *lane B*, 67% ammonium sulfate precipitate of the CFP; *lane C*, 67% ammonium-sulfate soluble material of the CFP; *CYT*, unfractionated cytosol; *lane D*, 29% ammonium sulfate precipitate of the cytosol; *lane E*, 44% ammonium sulfate precipitate of the cytosol; and *lane F*, 44% ammonium-sulfate soluble material of the cytosol. (**B**) Silver-stained SDS-PAGE of strong AIEX chromatography fractions using the 67% ammonium-sulfate precipitate of the CFP as an example. *Lane B*, unfractionated material; *lanes I-XIII* are the sequential AIEX fractions. (**C**) Silver-stained SDS-PAGE of RP chromatography fractions using the B-VI AIEX fraction of the 67% ammonium sulfate precipitate of the CFP and the F-XII AIEX fraction of the 44% ammonium-sulfate soluble material of the cytosol as examples. *Lanes B-VI and F-XII*, unfractionated material; *lanes 1-15*, the sequential RP fractions.

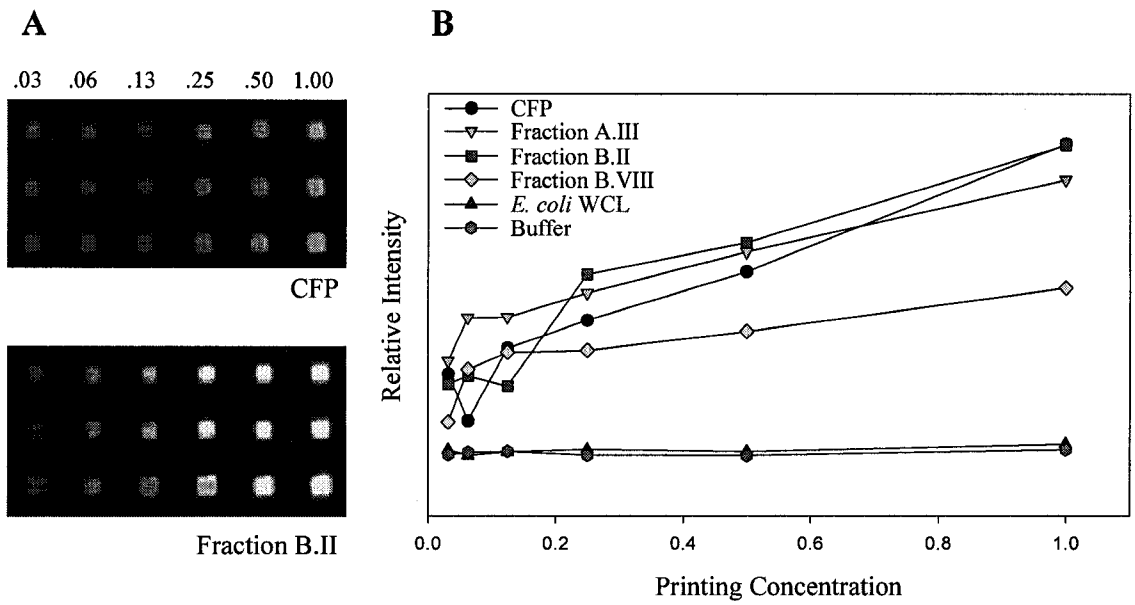


Fig. 3.2 Protein microarray printing optimization. (**A**) Examples of spotted dilution series of protein fractions and controls probed with TB patient sera and Cy3-labeled anti-human IgG. Spot size of 100 μM with 275 μM spacing. (**B**) The fluorescent intensities were quantitated and plotted against the printed protein concentrations. Fraction (X.Y) designation: X = Ammonium Sulfate cut A-F, Y = AIEX fraction (13 sequential elution fractions, Roman Numeral I to XIII).

3.3.2 Validation of the protein microarray format

Each multidimensional chromatography fraction, as well as intermediate fractions and recombinant proteins were printed to nitrocellulose microarray slides. Sera from PPD+ healthy (n=12), noncavitary TB (n=9), cavitary TB (n=11), and HIV+TB+ (n=10) individuals were probed against the microarrays in triplicate, and for each slide the NAvSNR (see Materials and Methods) was calculated for each protein or protein fraction. The integrity of the protein microarray and validation of this platform was determined by assessing the reactivity of TB patients' sera with selected protein fractions and purified proteins spotted on the microarray slides as controls, and comparing these data to published results obtained by plate ELISA (16, 17). To demonstrate specificity,

the reactivity of TB patients' sera to unfractionated CFP was compared to that of *E. coli* WCL and FAST® print buffer. The CFP is known to contain numerous B cell antigens (7), and as expected, cavitary TB patient sera displayed significantly greater reactivity to CFP than to either buffer alone (p-value = 0.001) or *E. coli* WCL (p-value = 0.002) (Fig. 3.3A). Although patient-to-patient variability in reactivity to unfractionated CFP was observed, each individual cavitary TB patient's serum recognized CFP more strongly than they did buffer or *E. coli* WCL.

A second validation control was performed with purified native 38-kDa protein PstS1/Rv0934, a previously characterized B cell antigen (7). Reactivity to PstS1 was compared among the four patient groups used in this study (Fig. 3.3B). Sera from TB patients showed greater reactivity than that from PPD+ healthy individuals. Furthermore, when evaluated using the mean NAvSNR of PPD+ healthy controls plus three times the standard deviation (SD) as a cutoff, the number of patients' sera demonstrating a positive response to PstS1 was greatest for cavitary TB (72%), followed by noncavitary TB (33%), and HIV+TB+ (30%) patients. These results concurred with data previously obtained via traditional ELISA platforms (16, 17). The PstS1 was also printed at multiple concentrations and a dose-dependent antibody response was observed (data not shown).

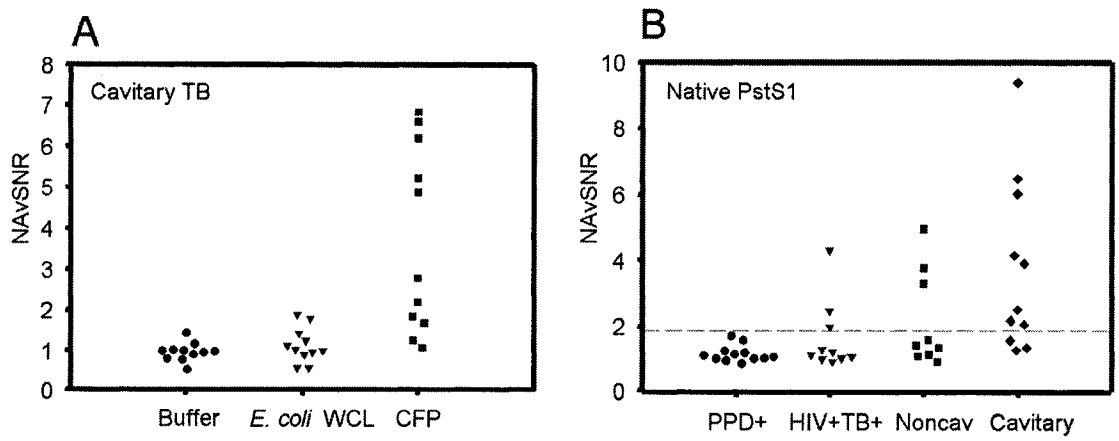


Fig. 3.3 Validation of protein microarray integrity. (**A**) Reactivity of cavitary TB patients' sera against buffer (●), *E. coli* WCL (▼), and *Mtb* CFP (■). (**B**) Reactivity of PPD+ healthy controls (●), HIV+TB+ (▼), noncavitary-TB (■), and cavitary-TB (♦) patients' sera against spotted native, purified 38-kDa antigen/PstS1. The patient-averaged NAvSNR +3X S.D. was obtained by use of sera from healthy PPD+ control subjects, shown as the horizontal dashed line, and was used as the cutoff to determine positive reactivity. *Noncav*, noncavitary.

3.3.3 TB disease states react to a defined group of antigens

A qualitative visual inspection of the probed microarray slides revealed overall differences in levels of reactivity based on disease state. (Fig. 3.4) Briefly, probing slides with cavitary TB sera resulted in the largest numbers of reactive spots while control PPD+ sera resulted in the fewest. Noncavitary TB sera displayed variable levels of reactivity from patient to patient, and HIV+TB+ sera generally displayed weak reactivity except for a distinct cluster of spots. To evaluate patterns of serological reactivity in a quantitative manner, the NAvSNR values of each fraction were averaged for all patients within a disease state, and expressed as a ratio over the corresponding fraction's averaged PPD+ healthy control NAvSNR (Fig. 3.5A). One of the more pronounced differences was observed when comparing HIV+TB+ patients to cavitary or noncavitary TB patients. The HIV+TB+ sera generally displayed poor reactivity, except for a distinct cluster of

CFP fractions that were not well recognized by sera from other disease states. A more general trend common to all three disease states was the increased reactivity of fractions originating from the CF (fractions 1-525) versus those derived from the cytosol (fractions 526-960).

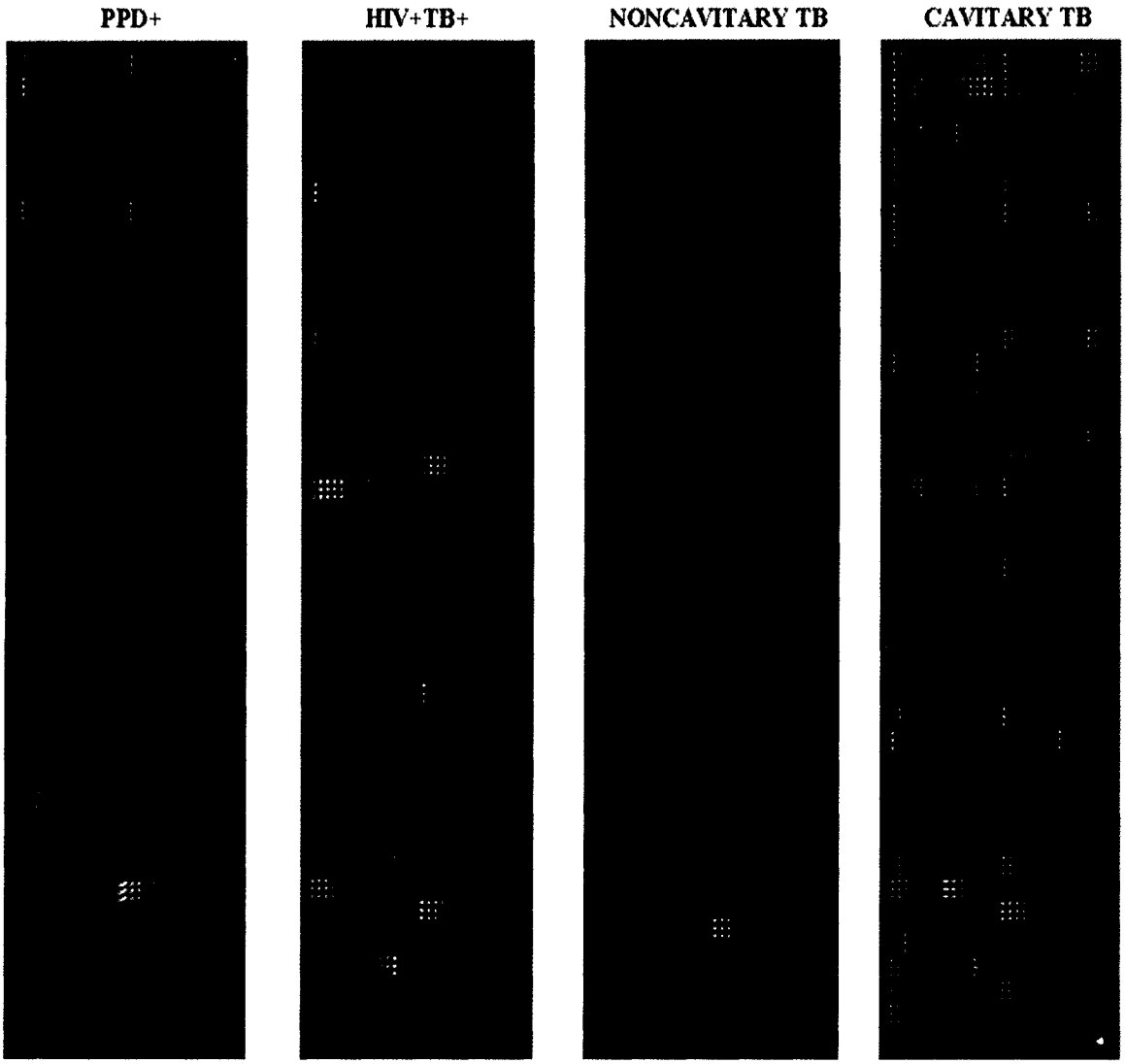


Fig. 3.4 Comparison of seroreactivity profiles across TB disease states. Shown are representative arrays probed with individual patient's serum and visualized with Cy3-labelled anti-human IgG.

To simplify patterns of antigen reactivity and to further assess similarities and differences in antigen recognition between disease states, a cutoff value of the mean NAvSNR of PPD+ healthy controls plus three SD was established for each fraction. Fractions displaying NAvSNR values greater than the cutoff value in 40% or more of the individuals in a disease state were selected and organized in a Venn diagram (Fig. 3.5B). Using these criteria, 145 of the 960 fractions (15%) were identified as having significant serological reactivity. As shown in Figure 3.5B, cavitary TB patients recognized 126 fractions, while noncavitary TB patients showed significant reactivity to 59 fractions, of which 55 were also recognized by cavitary TB patients. The remaining four fractions were unique to noncavitary TB patients. The pattern of fractions recognized by HIV+TB+ patients was less complex than that of either cavitary or noncavitary patients, and the overlap with these other two patient groups was minimal (Fig. 3.5B). HIV+TB+ patients recognized 20 fractions, of which five overlapped with those recognized by cavitary TB patients, and of this latter group two also reacted to noncavitary TB patients' sera. Interestingly, the 15 fractions that demonstrated significant reactivity to only HIV+TB+ patients' sera possessed very similar separation characteristics: 1) they originated from the CFP and did not precipitate with 67% ammonium sulfate, and 2) they bound strongly to the AIEX column, but weakly to the RP column. In Figure 3.5A these 15 fractions correspond to the cluster of CFP fractions that showed enhanced reactivity with HIV+TB+ patients' sera. To confirm that the antigen(s) in these fractions were specific to TB rather than HIV infection, the reactivity of three of the 15 fractions was assessed with HIV+TB- patients' sera, and no antibody responses significantly greater than that of PPD+ healthy controls were observed (data not shown).

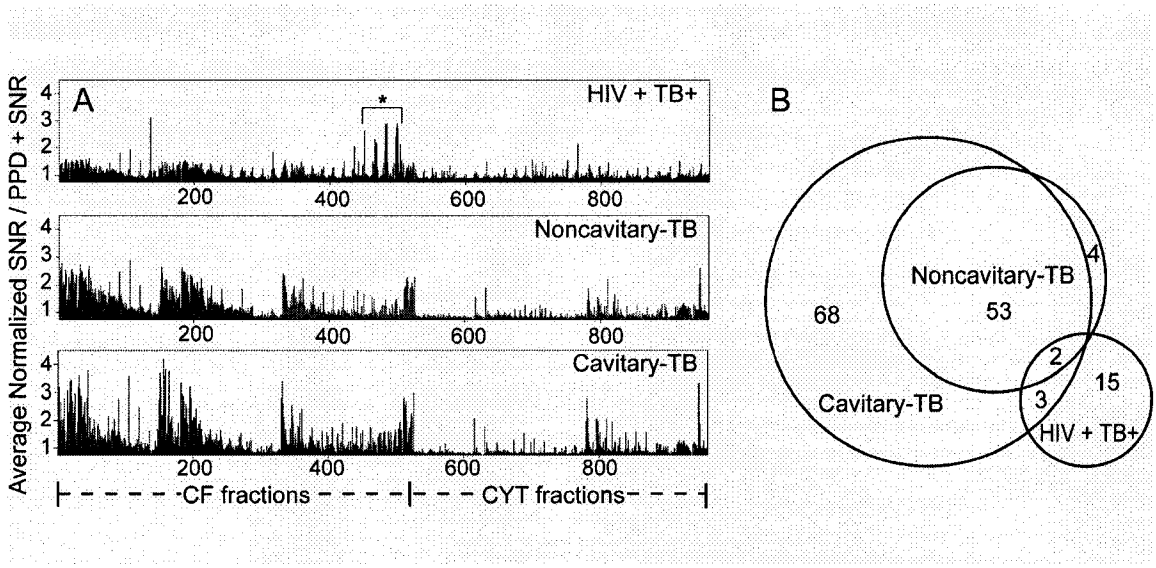


Fig. 3.5 Global Analysis of HIV+TB+, noncavitary TB, and cavitary TB patient reactivity against all 960 protein fractions. (**A**) The NAvSNR of each fraction was averaged for all patients in each disease state, and expressed as a ratio over the averaged healthy PPD+ NAvSNR of the corresponding fraction. CF fractions, 1-525; cytosolic fractions, 526-960. (**B**) A Venn diagram displaying the number of fractions with significant reactivity to sera from each disease state and the relatedness of these serological responses. The fractions included in the Venn diagram were recognized by ≥ 40% of patients. The *asterisk* indicates a unique set of fractions recognized by HIV+TB+ patients. *CYT*, cytosol.

3.3.4 Antigen identification of reactive fractions confirmed the seroreactivity to known antigens

To identify individual antigens with the greatest utility in a serodiagnostic assay additional stringency was applied to the array data, further reducing the number of fractions demonstrating significant reactivity from 145 to 105. This was achieved by restricting analyses of the cavitary TB specific fractions to those that yielded significant serological reactivity (≥ 3X SD above PPD+ mean) with 55% (6 of 11) or greater of cavitary TB patients.

It was also recognized that LAM, a well-characterized B cell antigen (13), would be present in some of the multidimensional fractions and was expected to be serologically

dominant. Thus, microarray slides were probed with the CS-35 MAb specific for LAM (Fig. 3.6). Of the 105-targeted fractions 24 were found to contain LAM. The serological dominance of LAM was confirmed with conventional one-dimensional immunoblotting using pooled TB patients' sera (Fig. 3.6). It is interesting to note that the LAM containing fractions consistently yielded microarray spots with the highest fluorescent intensities when probed with patients' sera. Most of the PPD+ healthy control patients' sera also reacted against these fractions. However, the reactivity to TB patients' sera was significantly stronger than that of the healthy PPD+ control sera. (Table 3.1) The twenty-four LAM-containing fractions were excluded from further antigen analyses.

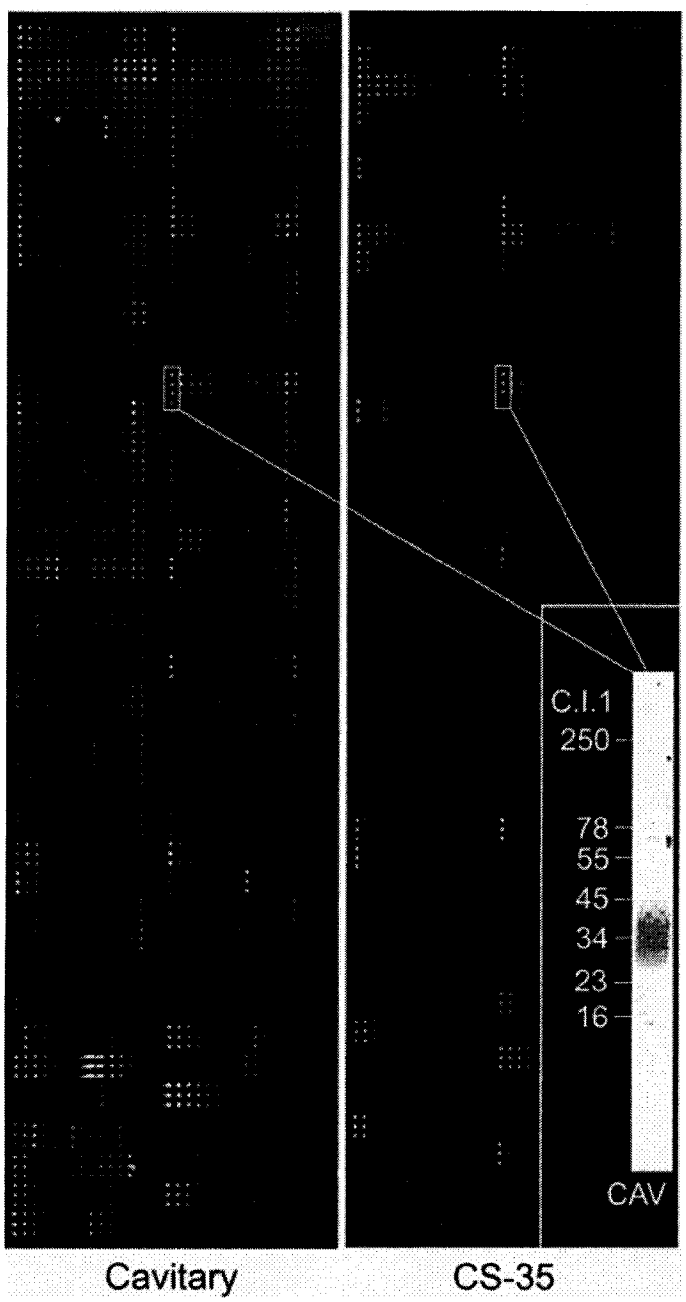


Fig. 3.6 Identification of reactive fractions containing lipoarabinomannan as the serodominant antigen. Shown are representative arrays probed with serum from a single cavitary TB patient, and the LAM-specific MAb CS-35. The inset is a Western blot of fraction C-I-1 probed with pooled cavitary-TB patients' sera..

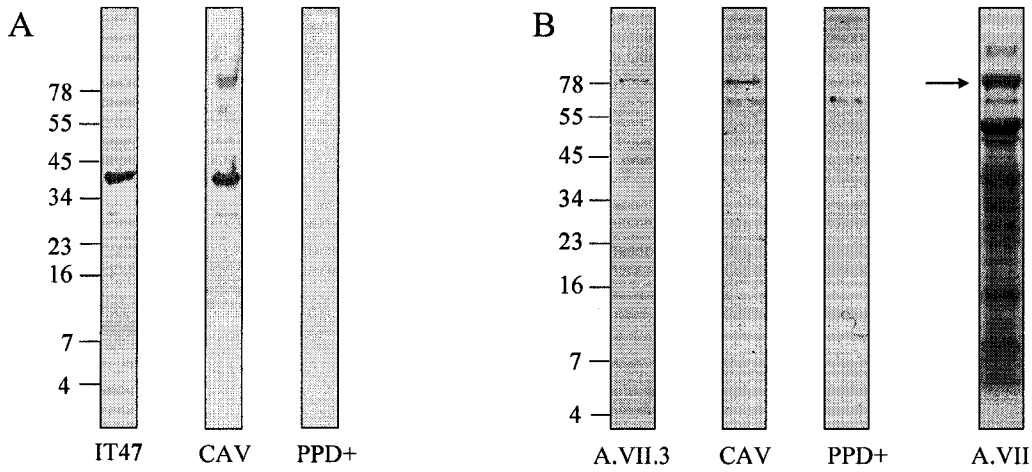
Table 3.1 Patient reactivity against LAM-containing fractions

Fractionation Conditions (X,Y,Z)[b]	LAM-Reactivity Ratio[a]			
	PPD+	HIV+ TB	Noncavitary TB	Cavitary TB
A-III-4	1.30	1.83	2.67	3.40
A-IV-1	3.08	3.49	5.71	6.31
A-IV-3	1.28	1.81	2.87	3.57
A-V-1	1.67	2.45	3.71	4.92
A-IX-1	1.41	2.48	3.33	3.94
A-X/XI-1	1.06	1.49	1.74	1.99
B-III-2	1.30	1.49	2.71	3.12
B-III-7	0.99	1.38	1.99	2.06
B-IV-1	2.30	2.61	4.33	5.72
B-I-2	1.23	1.53	2.43	3.23
C-III-1	1.18	1.41	2.14	2.19
C-IV-1	1.00	1.12	1.59	1.36
C-I-1	2.21	1.98	3.32	4.63
C-I-2	1.64	1.71	2.59	3.58
C-I-4	1.32	1.50	2.38	2.59
C-I-13	1.27	1.66	2.00	2.21
D-II/III/IV-1	1.68	1.84	3.15	3.91
E-II-1	2.91	2.90	3.70	4.68
E-III-1	1.07	1.45	1.66	1.47
F-II-2	1.07	1.22	1.65	2.33
F-II-3	0.98	1.24	1.40	1.60
F-IV-1	1.00	1.32	1.80	1.59
F-I-1	1.29	1.33	2.17	3.28
F-I-2	1.27	1.45	2.73	3.28

[a] Key to fraction designation: *X* = Ammonium Sulfate cut A-F (see Materials and Methods), *Y* = AIEX fraction (13 sequential elution fractions, Roman Numeral I to XIII), "/" denotes pooled fractions, *Z* = RP-HPLC fraction (sequential elutions 1 through 15).
[b] Disease state-averaged NAvSNR for the fraction / Disease state-averaged NAvSNR for all 960 fractions.

Molecular identities of the serologically active native proteins within the remaining fractions were obtained by two methods. The first of these utilized Western blot analyses with patients' sera compared alongside the reactivity to MAbs or polyclonal sera specific for five *Mtb* proteins (Fig. 3.7A). The second approach determined antigen composition of a fraction by Western blot analysis with patients' sera and identification

of the corresponding Coomassie-stained protein band via MS/MS (Fig. 3.7B and C). This combination of MS and antibody-based identification strategies resulted in an assigned antigen composition for 38 of the 81 remaining fractions.



C

Protein	Peptide Sequence	Observed m/z	Charge	Predicted MH+	Peptide AA Position
GlcB	VLYDFVNNEALPGTDIDPDSFWAGVDK	1500.37	2	2999.23	15-41
	VVADLTPQNQALLNAR	862.27	2	1723.95	42-57
	DELQAQIDKWHR	770.13	2	1539.68	58-69
	RVIEPIDMDAYR	739.63	2	1478.7	71-82
	WGSLYDALYGTDVIPETDGAEK	1200.63	2	2401.57	135-156
	WGSLYDALYGTDVIPETDGAEKGPTYNK	1531.16	2	3062.29	135-162
	NWLGLNK	844.30	1	844.98	290-296
	NWLGLNKGDLAAAVDKDGTAFLR	1223.27	2	2446.75	290-312
	GDLAAAVDKDGTAFLR	810.48	2	1620.79	297-312
	NYTAPGGGQFTLPGR	768.28	2	1536.67	319-333
	ASDVNGPLINSR	1243.42	1	1243.35	378-389
	TGSIYIVKPK	555.35	2	1106.34	390-399
	VEDVLGLPQNTMK	722.37	2	1444.68	416-428
	TGDEIHTSMEAGPMVR	866.01	2	1731.93	464-479
	SQPWILAYEDHNVDAGLAAGFSGR	1286.93	2	2575.78	485-508
	GM#WTM#TELM#ADM#VETK	969.47	2	1939.24	514-529
	RATIEQLLTIPLAK	783.99	2	1567.9	572-585
	ATIEQLLTIPLAK	705.97	2	1411.71	573-585
	VPDIHDVALMEDR	755.47	2	1510.7	622-634
	HGVITSADVR	1054.46	1	1055.17	650-659

Fig. 3.7 Antibody- and MS-based identification of antigens composing reactive fractions. (**A**) Western blot analyses of fraction C-II-11 with MAb IT-47, pooled cavitary-TB patients' sera, and pooled PPD+ healthy controls' sera (left to right) demonstrate the 38-kDa PstS1 protein as the dominant antigen in this fraction. (**B**) Fraction A-VII-3 was analyzed (left to right) by SDS-PAGE, silver staining, and Western blot with pooled sera from cavitary-TB and healthy PPD+ individuals to demonstrate an ~80-kDa protein as the dominant antigen. The corresponding coomassie-stained 80-kDa protein band (marked with →) of the parent AIEX fraction A-VII was digested with trypsin, and the (**C**) peptide sequences resolved by MS/MS identified the antigen as GlcB. *AA*, amino acid; *CAV*, cavitary; #, methionine oxidation.

Ten proteins previously found to be human B cell antigens (7) accounted for all or part of the serological activity of 26 fractions (Table 3.2). MAbs were used as the sole evidence to identify the presence of the 38-kDa PstS1 antigen, the 19-kDa lipoprotein antigen, and the 14-kDa HspX antigen in four reactive fractions. Specifically, the probing of a microarray slide with MAbs IT-23 and IT-47 identified PstS1 in fraction C-II-11. Western blot analyses with IT-47, cavitary TB sera, and PPD+ healthy control sera also revealed a single protein that reacted with IT-47 and cavitary TB patients' sera, but not sera of PPD+ healthy control individuals (Fig. 3.7A). In this same manner, the 19-kDa was found to be a reactive product of fractions B-I-7 and C-I-14, and HspX contributed to the reactivity of fraction B-V-9. It was also noted that fractions B-I-7 and C-I-14 each contained a second reactive product of 12 kDa and 45 kDa, respectively, and fraction B-V-9 contained two additional reactive proteins of 7 and 10 kDa. The identity of these unknown proteins was not determined due to insufficient protein quantities.

Table 3.2 Previously characterized antigens contained in serologically reactive fractions

Antigen(s)	Fractionation Conditions (X,Y,Z)[a]	Patient Reactivity[b] HIV+TB+	Noncavitary TB	Cavitary TB
38-kDa PstS1/Rv0934[c]	C-II-11	1/10 (10%)	1/9 (11%)	6/11 (55%)
45-kDa Apa/ModD /Rv1860[d]	A-II-11	2/10 (20%)	4/9 (44%)	4/11 (36%)
	A-III-9	2/10 (20%)	4/9 (44%)	5/11 (44%)
	A-III-11	2/10 (20%)	4/9 (44%)	7/11 (64%)
	B-III-10	2/10 (20%)	4/9 (44%)	4/11 (36%)
Ag85B/Rv1886c[e]	B-VI-13	3/10 (30%)	6/9 (67%)	6/11 (55%)
	B-VI-14	3/10 (30%)	3/9 (33%)	7/11 (64%)
	A-VI-15	2/10 (20%)	4/9 (44%)	5/11 (44%)
GlcB/Rv1837c[e]	A-VI-3	3/10 (30%)	4/9 (44%)	6/11 (55%)
	A-VI-4	2/10 (20%)	4/9 (44%)	6/11 (55%)
	A-VII-3	3/10 (30%)	4/9 (44%)	6/11 (55%)
	A-VIII-3	2/10 (20%)	4/9 (44%)	6/11 (55%)
	A-VIII-4	2/10 (20%)	4/9 (44%)	5/11 (44%)
LAM[c]	A-III-4	2/10 (10%)	4/9 (44%)	6/11 (55%)
	A-IV-1	1/10 (10%)	5/9 (56%)	6/11 (55%)
	A-IV-3	2/10 (20%)	4/9 (44%)	6/11 (55%)
	A-IX-1	3/10 (30%)	5/9 (56%)	6/11 (55%)
	A-V-1	3/10 (30%)	5/9 (56%)	6/11 (55%)
	A-X/XI-1	3/10 (30%)	4/9 (44%)	7/11 (64%)
	B-I-2	2/10 (20%)	4/9 (44%)	6/11 (55%)
	B-III-2	1/10 (10%)	4/9 (44%)	6/11 (55%)
	B-III-7	2/10 (20%)	3/9 (33%)	6/11 (55%)
	B-IV-1	0/10 (0%)	3/9 (33%)	7/11 (64%)
	C-I-1	0/10 (0%)	1/9 (11%)	6/11 (55%)
	C-I-2	0/10 (0%)	3/9 (33%)	6/11 (55%)
	C-I-4	2/10 (20%)	5/9 (56%)	6/11 (55%)
	C-I-13	2/10 (20%)	3/9 (33%)	5/11 (55%)
	C-III-1	2/10 (20%)	4/9 (44%)	6/11 (55%)
	C-IV-1	2/10 (20%)	4/9 (44%)	6/11 (55%)
	D-II/III/IV-1	0/10 (0%)	5/9 (56%)	7/11 (64%)
	E-II-1	0/10 (0%)	3/9 (33%)	6/11 (55%)
	E-III-1	2/10 (20%)	3/9 (33%)	6/11 (55%)
	F-I-1	0/10 (0%)	4/9 (44%)	7/11 (64%)
	F-I-2	2/10 (20%)	5/9 (56%)	6/11 (55%)
	F-II-2	2/10 (20%)	4/9 (44%)	7/11 (55%)
	F-II-3	2/10 (20%)	4/9 (44%)	7/11 (55%)
	F-IV-1	2/10 (20%)	3/9 (33%)	7/11 (64%)
Rv3881c[e]	A-VII-10	2/10 (20%)	3/9 (33%)	7/11 (64%)
	A-VII-11	2/10 (20%)	4/9 (44%)	7/11 (64%)
SecE2/Rv0379[e]	A-I-4	2/10 (20%)	4/9 (44%)	5/11 (44%)
	B-I-4	3/10 (30%)	5/9 (56%)	5/11 (44%)
	F-I-3	2/10 (20%)	3/9 (33%)	6/11 (55%)
	A-I-3	1/10 (10%)	4/9 (44%)	6/11 (55%)

Ag85A/Rv3804c & Ag85B/Rv1886c [e]	A-I-14	2/10 (20%)	3/9 (33%)	6/11 (55%)
	A-I-13	2/10 (20%)	4/9 (44%)	5/11 (44%)
	A-VI-12	2/10 (20%)	2/9 (22%)	6/11 (55%)
45 kDa & MPT64/Rv1980c [d]	A-III-10	2/10 (20%)	3/9 (33%)	6/11 (55%)
HspX/Rv2031c & 2 unknowns [c]	B-V-9	2/10 (20%)	0/9 (0%)	6/11 (55%)
19-kDa/Rv3763 & 1 unknown [c]	B-I-7	2/10 (20%)	4/9 (44%)	4/11 (36%)
	C-I-14	2/10 (20%)	3/9 (44%)	6/11 (55%)

[a] Key to fraction designation: X = Ammonium Sulfate cut A-F (see Materials and Methods), Y = AIEX fraction (13 sequential elution fractions, Roman Numeral I to XIII), "/" denotes pooled fractions, Z = RP-HPLC fraction (sequential elutions 1 through 15).

[b] Number (Percentage) of patient sera significantly reactive against fraction in question. Significant is > 3 X SD above PPD+ mean.

[c] Antigen identification based on reactivity to an antigen-specific MAb.

[d] Antigen identification based on reactivity to an antigen-specific MAb and MS/MS analyses.

[e] Antigen identification based on MS/MS analyses. Detailed MS/MS results are provided in Table 3.3.

The remaining seven previously described protein antigens (45 kDa Apa, 30 kDa Ag85A and B, GlcB, Rv3881c, SecE2, and MPT64) were identified by MS/MS analyses and some of these were confirmed by reactivity to MAbs (Table 3.2 and 3.3). For some fractions, after Western blot analyses with patients' sera a sufficient amount of material was not available for protein identification. Thus, protein identity was obtained by MS/MS analysis of an adjacent fraction possessing a reactive band at the same molecular weight. This was done to identify the 45 kDa Apa in one fraction, Ag85B in four fractions, Ag85A in two fractions, GlcB in four fractions, SecE2 in three fractions, and Rv3881c in one fraction (Table 3.2). The identification of four novel antigens (SodC, LppZ, BfrB, and TrxC) will be further described in Chapter 4.

Table 3.3 MS-based identification of previously identified antigens

Protein accession number: Rv1860 Molecular Weight (Da): 32702.8
Common Name(s): ModD, Apa, 45 kDa AA Coverage: 47% # of total spectra: 65 # of unique spectra: 12 # of unique peptides: 8
Fraction: A-III-9*

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	0	0	4.96		1719.98	2	3437.95
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	0	0	2.96		1719.97	2	3437.92
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	0	0	3.19		1720.04	2	3438.07
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	2.82	0.545	1.75		1720.34	2	3438.67
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	2.46	0.419	1.19		1720.60	2	3439.19
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	2.92	0.607	2.36		1721.09	2	3440.16
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	0	0	3.1		1721.73	2	3441.44
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	0	0	5.51		1146.99	3	3437.95
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	0	0	6.77		1147.00	3	3437.98
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	0	0	6.36		1147.06	3	3438.17
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	0	0	10.5		1147.03	3	3438.08
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	0	0	8.68		1147.07	3	3438.19
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	5.03	0.631	7		1147.11	3	3438.3
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	4.19	0.649	5.7		1147.25	3	3438.74
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	5.27	0.68	6.46		1147.25	3	3438.74
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	4.24	0.666	4.14		1147.37	3	3439.09
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	4.32	0.64	6.96		1147.43	3	3439.27
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	4.92	0.625	4.89		1147.40	3	3439.17
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	4.33	0.643	6		1147.62	3	3439.85
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	4.86	0.728	3.04		1147.86	3	3440.56
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	4.24	0.55	1.23		1148.00	3	3440.98
TTGDPPFPGQPPPVANDTR	K	I	1.81	0.493	-0.176		1964.83	1	1963.82
TTGDPPFPGQPPPVANDTR	K	I	2.42	0.591	0.585		982.60	2	1963.19
TTGDPPFPGQPPPVANDTR	K	I	2.28	0.548	1.51		982.71	2	1963.4
TTGDPPFPGQPPPVANDTR	K	I	2.62	0.475	3.77		982.79	2	1963.57
TTGDPPFPGQPPPVANDTR	K	I	3.57	0.543	7.96		982.78	2	1963.55
TTGDPPFPGQPPPVANDTR	K	I	2.02	0.386	0.328		982.99	2	1963.96
TTGDPPFPGQPPPVANDTR	K	I	3.83	0.57	0.482		983.04	2	1964.07
TTGDPPFPGQPPPVANDTR	K	I	3.27	0.567	1.77		983.29	2	1964.56
LGSDMGEFYMPYPGTR	R	I	2.34	0.353	0		1821.68	1	1820.67
LGSDMGEFYMPYPGTR	R	I	2.44	0.427	0		911.19	2	1820.37
LGSDMGEFYMPYPGTR	R	I	3.3	0.583	0		911.14	2	1820.26
LGSDMGEFYMPYPGTR	R	I	3.38	0.478	0		911.22	2	1820.43
LGSDMGEFYMPYPGTR	R	I	2.57	0.466	0		911.45	2	1820.88
LGSDMGEFYMPYPGTR	R	I	2.18	0.472	0		911.40	2	1820.78
LGSDMGEFYMPYPGTR	R	I	2.5	0.319	0	Oxidation (+16)	919.10	2	1836.18
LGSDMGEFYMPYPGTR	R	I	3.53	0.569	3	2 X Oxidation (+16)	926.97	2	1851.93
LGSDMGEFYMPYPGTR	R	I	3.19	0.551	2.06	2 X Oxidation (+16)	927.17	2	1852.32
INQETVSLDANGVSGSASYYEVK	R	F	4.87	0.718	4.09		1216.43	2	2430.85
INQETVSLDANGVSGSASYYEVK	R	F	5.97	0.71	3.62		1216.68	2	2431.34
INQETVSLDANGVSGSASYYEVK	R	F	4.9	0.75	6.89		1216.74	2	2431.46
INQETVSLDANGVSGSASYYEVK	R	F	6.14	0.754	3.3		1216.77	2	2431.53
INQETVSLDANGVSGSASYYEVK	R	F	5.32	0.703	4.64		1216.81	2	2431.6
INQETVSLDANGVSGSASYYEVK	R	F	5.16	0.713	1.46		1216.87	2	2431.72
INQETVSLDANGVSGSASYYEVK	R	F	5.48	0.737	3.51		1216.84	2	2431.67
INQETVSLDANGVSGSASYYEVK	R	F	5.66	0.736	3.48		1216.85	2	2431.68
INQETVSLDANGVSGSASYYEVK	R	F	4.12	0.636	2.57		1216.87	2	2431.72
INQETVSLDANGVSGSASYYEVK	R	F	5.54	0.717	4.68		1216.95	2	2431.88
INQETVSLDANGVSGSASYYEVK	R	F	5.58	0.688	5.68		1216.94	2	2431.86
INQETVSLDANGVSGSASYYEVK	R	F	5.54	0.732	4.96		1217.05	2	2432.09
INQETVSLDANGVSGSASYYEVK	R	F	4.9	0.719	5.04		1217.00	2	2431.99
INQETVSLDANGVSGSASYYEVK	R	F	4.76	0.722	4.47		1217.20	2	2432.38
INQETVSLDANGVSGSASYYEVK	R	F	5.5	0.661	2.38		1217.18	2	2432.34
INQETVSLDANGVSGSASYYEVK	R	F	5.27	0.698	1.85		1217.29	2	2432.57
INQETVSLDANGVSGSASYYEVK	R	F	5.67	0.684	5.21		1217.49	2	2432.96
WFVVWLGTANNPVDK	R	G	1.94	0.183	2.92		1745.77	1	1744.76
WFVVWLGTANNPVDKGAAK	R	A	4.73	0.525	6.1		1037.27	2	2072.52
FSDPSKPNGQIWTGVIGSPAANAPDAGPPQR	K	W	3.23	0.688	5		1567.46	2	3132.91
FSDPSKPNGQIWTGVIGSPAANAPDAGPPQR	K	W	4.15	0.727	0		1567.52	2	3133.02
FSDPSKPNGQIWTGVIGSPAANAPDAGPPQR	K	W	4.03	0.788	1.54		1568.01	2	3134

FSDPSKPNGQIWTGVIGSPAANAPDAGPPQR	K	W	6.01	0.694	3.38		1045.70	3	3134.08
LYASAEATDSK	K	A	1.95	0.372	2.46		1155.39	1	1154.38
LYASAEATDSK	K	A	2.16	0.494	3.27		1155.40	1	1154.39
LYASAEATDSK	K	A	1.17	0.441	2.59		1155.53	1	1154.52
LYASAEATDSK	K	A	2.49	0.692	3.96		1157.60	1	1156.59

Protein accession number: Rv1860 Molecular Weight (Da): 32702.8
Common Name(s): ModD, Apa, 45 kDa AA Coverage: 17% # of total spectra: 6 # of unique spectra: 3 # of unique peptides: 2
Fraction: A-III-11

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	3.09	0.633	2.82		1720.60	2	3439.19
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	2.77	0.527	3.38		1147.40	3	3439.19
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	1.28	0.471	0		1147.45	3	3439.33
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	2.74	0.495	0.959		1147.62	3	3439.85
INQETVSLDANGVSGSASYYEVK	R	F	2.89	0.636	5.21		1216.68	2	2431.34
INQETVSLDANGVSGSASYYEVK	R	F	4.08	0.707	-0.279		1216.81	2	2431.61

Protein accession number: Rv1860 Molecular Weight (Da): 32702.8
Common Name(s): ModD, Apa, 45 kDa AA Coverage: 22% # of total spectra: 21 # of unique spectra: 3 # of unique peptides: 3
Fraction: B-III-10

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
LGSDMGEFYMPYPGTR	R	I	2.85	0.632	0		910.99	2	1819.96
LGSDMGEFYMPYPGTR	R	I	3.28	0.493	0		911.08	2	1820.14
LGSDMGEFYMPYPGTR	R	I	2.59	0.424	0	Oxidation (+16)	918.96	2	1835.9
LGSDMGEFYMPYPGTR	R	I	2.24	0.459	0	Oxidation (+16)	919.52	2	1837.02
INQETVSLDANGVSGSASYYEVK	R	F	4.23	0.625	0		1216.41	2	2430.8
INQETVSLDANGVSGSASYYEVK	R	F	2.68	0.576	2.8		1216.45	2	2430.89
INQETVSLDANGVSGSASYYEVK	R	F	2.46	0.558	1.89		1216.37	2	2430.73
INQETVSLDANGVSGSASYYEVK	R	F	1.89	0.612	-0.732		1216.52	2	2431.02
INQETVSLDANGVSGSASYYEVK	R	F	2.5	0.632	2.96		1216.58	2	2431.14
INQETVSLDANGVSGSASYYEVK	R	F	4.71	0.68	4.05		1216.59	2	2431.16
INQETVSLDANGVSGSASYYEVK	R	F	4.94	0.719	4.82		1216.61	2	2431.21
INQETVSLDANGVSGSASYYEVK	R	F	3.23	0.596	1.39		1216.66	2	2431.3
INQETVSLDANGVSGSASYYEVK	R	F	4.56	0.659	0.585		1216.72	2	2431.43
INQETVSLDANGVSGSASYYEVK	R	F	4.41	0.758	2.47		1216.88	2	2431.74
INQETVSLDANGVSGSASYYEVK	R	F	4.88	0.657	4.42		1216.99	2	2431.97
INQETVSLDANGVSGSASYYEVK	R	F	2.62	0.499	1.54		1217.13	2	2432.24
INQETVSLDANGVSGSASYYEVK	R	F	5.63	0.721	3.42		1217.20	2	2432.38
INQETVSLDANGVSGSASYYEVK	R	F	1.41	0.448	1.37		1217.26	2	2432.5
INQETVSLDANGVSGSASYYEVK	R	F	3.84	0.571	1.77		1217.27	2	2432.52
INQETVSLDANGVSGSASYYEVK	R	F	4.33	0.715	5		1217.54	2	2433.07
FSDPSKPNGQIWTGVIGSPAANAPDAGPPQR	K	W	3.26	0.576	1.46		1045.78	3	3134.32

Protein accession number: Rv1860 Molecular Weight (Da): 32702.8
Common Name(s): ModD, Apa, 45 kDa AA Coverage: 36% # of total spectra: 30 # of unique spectra: 7 # of unique peptides: 6
Fraction: A-III-10

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	2.07	0.47	0		1720.33	2	3438.64
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	1.73	0.427	4.47		1147.11	3	3438.31
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	2.83	0.53	5.38		1147.15	3	3438.44
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	2.51	0.471	0.495		1147.12	3	3438.33
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	3.62	0.576	9.14		1147.29	3	3438.86
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	4.84	0.573	5.14		1147.53	3	3439.57
IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	R	T	0.922	0.163	2.5		1147.93	3	3440.76
TTGDPPFPGQPPPVANDTR	K	I	3.2	0.547	1.28		982.83	2	1963.65
LGSDMGEFYMPYPGTR	R	I	3.12	0.576	0		911.03	2	1820.04
LGSDMGEFYMPYPGTR	R	I	1.32	0.339	2.34	2 X Oxidation (+16)	927.35	2	1852.69
INQETVSLDANGVSGSASYYEVK	R	F	1.82	0.448	3.12		1216.12	2	2430.23
INQETVSLDANGVSGSASYYEVK	R	F	1.88	0.596	0.377		1216.35	2	2430.69
INQETVSLDANGVSGSASYYEVK	R	F	4.47	0.675	5.64		1216.53	2	2431.05
INQETVSLDANGVSGSASYYEVK	R	F	1.95	0.471	3.1		1216.50	2	2430.98
INQETVSLDANGVSGSASYYEVK	R	F	1.48	0.462	0		1216.59	2	2431.17
INQETVSLDANGVSGSASYYEVK	R	F	4.71	0.752	3.96		1216.68	2	2431.34

INQETVSLDANGVSGSASYYEVK	R	F	5.38	0.732	5.52		1216.66	2	2431.3
INQETVSLDANGVSGSASYYEVK	R	F	5.17	0.664	0		1216.69	2	2431.37
INQETVSLDANGVSGSASYYEVK	R	F	4.15	0.696	6.77		1216.77	2	2431.52
INQETVSLDANGVSGSASYYEVK	R	F	4.59	0.735	0		1216.83	2	2431.65
INQETVSLDANGVSGSASYYEVK	R	F	2.27	0.6	-1.11		1216.91	2	2431.81
INQETVSLDANGVSGSASYYEVK	R	F	5.09	0.735	6.31		1216.93	2	2431.84
INQETVSLDANGVSGSASYYEVK	R	F	4.97	0.62	1.42		1216.98	2	2431.95
INQETVSLDANGVSGSASYYEVK	R	F	4.73	0.672	4.47		1217.12	2	2432.22
INQETVSLDANGVSGSASYYEVK	R	F	4.65	0.664	6.03		1217.08	2	2432.15
INQETVSLDANGVSGSASYYEVK	R	F	4.73	0.642	3.03		1217.13	2	2432.25
INQETVSLDANGVSGSASYYEVK	R	F	5.15	0.617	6.48		1217.43	2	2432.84
INQETVSLDANGVSGSASYYEVK	R	F	0	0	4.22		1218.17	2	2434.32
WFVVWLGTANNPVDK	R	G	2.46	0.261	2.8		873.34	2	1744.66
LYASAEATDSK	K	A	1.9	0.536	-0.114		1155.51	1	1154.5

Protein accession number: Rv3804c Molecular Weight (Da): 35668.3
Common Name(s): FbpA, Ag85A AA Coverage: 26% # of total spectra: 13 # of unique spectra: 7 # of unique peptides: 6
Fraction: A-I-13

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
ALGATPNTGPAPQGA	R	-	2.97	0.62	4.6		1323.09	1	1322.08
ALGATPNTGPAPQGA	R	-	1.53	0.283	4.39		661.94	2	1321.86
WETFLTSELPGWLQANR	K	H	4.64	0.682	5.8		1025.04	2	2048.07
NDPLLNVGK	R	L	1.24	0.528	2		969.39	1	968.38
VWVYCGNGKPSDLGGNNLPAK	R	F	2.38	0.619	-0.556	Propionamide (+71)	1130.97	2	2259.92
VWVYCGNGKPSDLGGNNLPAK	R	F	5.22	0.732	3.75	Propionamide (+71)	1131.14	2	2260.27
FLEGFVR	K	T	1.82	0.324	1.22		434.52	2	867.03
VQFQSGGANSPALYLLDGLR	K	A	2.33	0.419	0.495		1053.55	2	2105.09
VQFQSGGANSPALYLLDGLR	K	A	2.73	0.478	1.92		1053.69	2	2105.36
VQFQSGGANSPALYLLDGLR	K	A	4.21	0.512	10.5		1053.66	2	2105.31
VQFQSGGANSPALYLLDGLR	K	A	5.57	0.665	9.17		1054.02	2	2106.03
VQFQSGGANSPALYLLDGLR	K	A	5.49	0.608	5.96		1054.35	2	2106.69
VQFQSGGANSPALYLLDGLR	K	A	2.33	0.614	4.47		1054.57	2	2107.13

Protein accession number: Rv1886c Molecular Weight (Da): 34563.6
Common Name(s): FbpB, Ag85B AA Coverage: 25% # of total spectra: 8 # of unique spectra: 4 # of unique peptides: 4
Fraction: A-I-13

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
VQFQSGGNNSPAVYLLDGLR	K	A	3.87	0.69	9.04		1068.27	2	2134.52
VQFQSGGNNSPAVYLLDGLR	K	A	3.57	0.557	2.11		1068.63	2	2135.24
VQFQSGGNNSPAVYLLDGLR	K	A	0	0	4.7		1070.11	2	2138.21
LWVYCGNGTPNELGGANIPAEFLENFVR	R	S	2.57	0.461	1.46	Propionamide (+71)	1577.08	2	3152.15
LWVYCGNGTPNELGGANIPAEFLENFVR	R	S	1.87	0.575	3.19	Propionamide (+71)	1577.28	2	3152.54
PGLPVEYLQVPSPSMGR	R	D	4.04	0.697	0		914.57	2	1827.13
PGLPVEYLQVPSPSMGR	R	D	4.33	0.595	0	Oxidation (+16)	922.41	2	1842.8
WETFLTSELPQWLSANR	K	A	3.12	0.471	4.17		1039.72	2	2077.43

Protein accession number: Rv1886c Molecular Weight (Da): 34563.6
Common Name(s): FbpB, Ag85B AA Coverage: 30% # of total spectra: 11 # of unique spectra: 6 # of unique peptides: 5
Fraction: A-VI-12

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
VQFQSGGNNSPAVYLLDGLR	K	A	4.02	0.61	6.48		1068.19	2	2134.37
VQFQSGGNNSPAVYLLDGLR	K	A	4.71	0.641	5.11		1068.25	2	2134.48
VQFQSGGNNSPAVYLLDGLR	K	A	6.08	0.69	4.48		1069.13	2	2136.24
AADMWGPSSDPAWER	K	N	2.73	0.587	0		838.48	2	1674.95
AADMWGPSSDPAWER	K	N	3.26	0.479	0		838.67	2	1675.33
LWVYCGNGTPNELGGANIPAEFLENFVR	R	S	3.2	0.385	3.62	Propionamide (+71)	1576.86	2	3151.7
LWVYCGNGTPNELGGANIPAEFLENFVR	R	S	3.4	0.596	2.54	Propionamide (+71)	1577.28	2	3152.55
LWVYCGNGTPNELGGANIPAEFLENFVR	R	S	4.96	0.657	3.1	Propionamide (+71)	1052.21	3	3153.62
WETFLTSELPQWLSANR	K	A	3.59	0.611	5.7		1039.73	2	2077.45
PGLPVEYLQVPSPSMGR	R	D	4.37	0.729	0		914.23	2	1826.44
PGLPVEYLQVPSPSMGR	R	D	3.74	0.603	0	Oxidation (+16)	922.02	2	1842.03

Protein accession number: Rv1837c | Molecular Weight (Da): 80386.3
Common Name(s): GlcB, 81/88 kDa | AA Coverage: 35% | # of total spectra: 45 | # of unique spectra: 23 | # of unique peptides: 20
Fraction: A-VII-3

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
TGSIYIVKPK	R	M	2.81	0.484	3.03		553.35	2	1104.68
ASDVNGPLINSR	K	T	1.83	0.423	0.538		1243.42	1	1242.41
ASDVNGPLINSR	K	T	3.04	0.601	0.721		622.72	2	1243.43
WGSLYDALYGTDVIPETDGAEKGPTYNK	R	V	3.53	0.759	9.29		1531.16	2	3060.31
WGSLYDALYGTDVIPETDGAEKGPTYNK	R	V	1.89	0.393	2.68		1531.36	2	3060.7
WGSLYDALYGTDVIPETDGAEKGPTYNK	R	V	5.03	0.646	13.2		1020.94	3	3059.8
NWLGLNK	R	G	1.66	0.398	2.07		844.30	1	843.29
NWLGLNK	R	G	1.66	0.433	0		844.86	1	843.85
VPDIHDVALMEDR	K	A	3.84	0.642	0		755.47	2	1508.93
VPDIHDVALMEDR	K	A	3.29	0.476	6.82	Oxidation (+16)	763.24	2	1524.47
WGSLYDALYGTDVIPETDGAEK	R	G	1.96	0.484	3.47		1200.63	2	2399.24
WGSLYDALYGTDVIPETDGAEK	R	G	4.03	0.689	5.7		1200.73	2	2399.45
HGVITSADVR	R	A	1.63	0.456	1.57		1054.46	1	1053.45
HGVITSADVR	R	A	1.83	0.413	1.96		527.87	2	1053.73
HGVITSADVR	R	A	2.93	0.538	2.26		527.93	2	1053.84
NYTAPGGGQFTLPGR	R	S	3.08	0.591	5.7		768.28	2	1534.54
NYTAPGGGQFTLPGR	R	S	1.99	0.577	1.89		768.42	2	1534.82
NYTAPGGGQFTLPGR	R	S	2.96	0.565	3.06		768.87	2	1535.72
RATIEQLLTIPLAK	R	E	3.77	0.471	2.05		783.99	2	1565.97
VVADLTPQNQALLNAR	K	D	3.29	0.615	3.8		862.27	2	1722.52
VVADLTPQNQALLNAR	K	D	4.1	0.635	3.09		862.59	2	1723.17
VVADLTPQNQALLNAR	K	D	3.44	0.702	4.59		862.66	2	1723.3
VLYDFVNNEALPGTDIDPDSFWAGVDK	R	V	3.09	0.679	1.24		1500.37	2	2998.73
VLYDFVNNEALPGTDIDPDSFWAGVDK	R	V	4.19	0.67	4.57		1500.70	2	2999.39
NWLGLNKGDLAAAVDKDGTAFLR	R	V	3.8	0.641	7.37		1223.27	2	2444.52
SQPWILAYEDHNVDAGLAAGFSGR	K	A	0	0	3.19		1286.93	2	2571.85
SQPWILAYEDHNVDAGLAAGFSGR	K	A	0	0	2.92		1287.50	2	2572.98
SQPWILAYEDHNVDAGLAAGFSGR	K	A	0	0	3.24		1287.59	2	2573.16
SQPWILAYEDHNVDAGLAAGFSGR	K	A	1.86	0.525	0.292		1287.79	2	2573.56
SQPWILAYEDHNVDAGLAAGFSGR	K	A	2.54	0.58	0		1288.34	2	2574.67
SQPWILAYEDHNVDAGLAAGFSGR	K	A	3.32	0.62	4.62		1288.81	2	2575.61
RVIEPIDMDAYR	R	Q	4.34	0.523	0		739.63	2	1477.25
RVIEPIDMDAYR	R	Q	2.15	0.558	1.43	Oxidation (+16)	747.51	2	1493
ATIEQLLTIPLAK	R	E	3.27	0.422	4.72		705.97	2	1409.92
GDLAAAVDKDGTAFLR	K	V	2.6	0.49	5.02		810.48	2	1618.95
GDLAAAVDKDGTAFLR	K	V	2.92	0.508	4.32		810.46	2	1618.91
TGDEIHTSMEAGPMVR	R	K	3.47	0.327	0		866.01	2	1730
TGDEIHTSMEAGPMVR	R	K	3.21	0.596	0	Oxidation (+16)	874.08	2	1746.14
TGDEIHTSMEAGPMVR	R	K	2.22	0.34	1.31	2 X Oxidation (+16)	882.04	2	1762.06
DELQAQIDKWHR	R	R	3.93	0.54	3.68		770.13	2	1538.25
VEDVLGLPQNTMK	R	I	3.01	0.411	0		722.37	2	1442.72
VEDVLGLPQNTMK	R	I	2.64	0.576	1.09	Oxidation (+16)	730.54	2	1459.07
GMWTMTELMADMVETK	K	I	2.31	0.469	0	Oxidation (+16)	945.78	2	1889.54
GMWTMTELMADMVETK	K	I	3	0.607	3.66	4 X Oxidation (+16)	969.47	2	1936.92
GMWTMTELMADMVETK	K	I	3.11	0.424	3.75	4 X Oxidation (+16)	969.68	2	1937.34

Protein accession number: Rv1980c | Molecular Weight (Da): 24805.3
Common Name(s): MPT64 | AA Coverage: 26% | # of total spectra: 4 | # of unique spectra: 3 | # of unique peptides: 2
Fraction: A-III-10

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
GTDTGQACQIQMSDPAYNINISLPSYYPDQK	K	S	1.37	0.472	0	Propionamide (+71)	1746.49	2	3490.97
KPITYDTLWQADTDPLPVVFPIVQGELSK	R	Q	0	0	10.2		1636.08	2	3270.15
KPITYDTLWQADTDPLPVVFPIVQGELSK	R	Q	4.45	0.588	5.26		1091.22	3	3270.65
KPITYDTLWQADTDPLPVVFPIVQGELSK	R	Q	4.75	0.567	3.48		1091.47	3	3271.39

Protein accession number: Rv3881c | Molecular Weight (Da): 47575.7
Common Name(s): | AA Coverage: 27% | # of total spectra: 19 | # of unique spectra: 8 | # of unique peptides: 8
Fraction: A-VII-10

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
VATAGEPNFMDLK	R	E	2.63	0.494	0		697.10	2	1392.18
VATAGEPNFMDLK	R	E	2.46	0.53	3.82	Oxidation (+16)	704.94	2	1407.87
VATAGEPNFMDLK	R	E	2.26	0.479	3.35	Oxidation (+16)	704.94	2	1407.86
REHPTYEDIVGLER	R	L	3.84	0.641	1.44		857.66	2	1713.31
REHPTYEDIVGLER	R	L	4.5	0.567	4.28		858.61	2	1715.2
AALEPVNPPKPPPAIK	K	I	2.23	0.543	4.59		820.01	2	1638
AALEPVNPPKPPPAIK	K	I	2.45	0.628	3.42		820.39	2	1638.76
ANEVEAPMADPPTDVPITPCELTAAK	R	N	2.55	0.615	0		1341.21	2	2680.41
ANEVEAPMADPPTDVPITPCELTAAK	R	N	3.32	0.441	0	Propionamide (+71)	1376.74	2	2751.46
ANEVEAPMADPPTDVPITPCELTAAK	R	N	0	0	4.85	Propionamide (+71), Oxidation (+16)	1384.26	2	2766.51
ANEVEAPMADPPTDVPITPCELTAAK	R	N	1.59	0.328	5.03	Oxidation (+16), Propionamide (+71)	1384.36	2	2766.7
DQILPVYAEYQQR	R	S	3.37	0.614	8.68		812.23	2	1622.45
DQILPVYAEYQQR	R	S	3.59	0.675	10.1		812.31	2	1622.61
QWILHMAK	R	L	1.9	0.507	0		513.89	2	1025.76
NAAQQLVLSADNMR	K	E	4.31	0.632	0		765.93	2	1529.85
NAAQQLVLSADNMR	K	E	4.56	0.654	0		765.96	2	1529.91
NAAQQLVLSADNMR	K	E	3.08	0.66	0		766.38	2	1530.74
NAAQQLVLSADNMR	K	E	4.02	0.678	5.96	Oxidation (+16)	774.18	2	1546.34
GFDNWEGDAATACEASLDQQR	R	Q	3.82	0.617	6.85	Propionamide (+71)	1178.31	2	2354.61

Protein accession number: Rv0379 | Molecular Weight (Da): 7948.1
Common Name(s): SecE2 | Total AA Coverage: 73%
Fraction: B-1-4

Trypsin digestion: AA Coverage: 31% | # of total spectra: 2 | # of unique spectra: 1 | # of unique peptides: 1

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
VIDIIGTSPTSWEQAAAEAVQR	K	A	3.26	0.504	6.57		1171.56	2	2341.1
VIDIIGTSPTSWEQAAAEAVQR	K	A	2.97	0.583	5.59		1171.74	2	2341.46

Chymotrypsin digestion: AA Coverage: 30% | # of total spectra: 6 | # of unique spectra: 3 | # of unique peptides: 2

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
KMRPAQPR	F	-	1.92	0.313	0		492.46	2	982.9
KVIDIIGTSPTSW	Y	E	0.965	0.0505	4.82		1416.61	1	1415.6
KVIDIIGTSPTSW	Y	E	2.72	0.435	0.108		1418.59	1	1417.58
KVIDIIGTSPTSW	Y	E	0	0	3.8		708.24	2	1414.47
KVIDIIGTSPTSW	Y	E	3.15	0.493	6.92		708.95	2	1415.88
KVIDIIGTSPTSW	Y	E	3.59	0.43	4.85		709.06	2	1416.1

Glu-C digestion: AA Coverage: 41% | # of total spectra: 10 | # of unique spectra: 3 | # of unique peptides: 2

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
VSFKMRPAQPR	E	-	2.01	0.576	0		1318.60	1	1317.59
VSFKMRPAQPR	E	-	2.02	0.309	0		659.14	2	1316.26
QDMAVDSAGKITYRIKLE	E	V	1.76	0.442	2.57		1019.45	2	2036.89
QDMAVDSAGKITYRIKLE	E	V	1.69	0.449	3.33		1019.53	2	2037.04
QDMAVDSAGKITYRIKLE	E	V	1.56	0.427	3.96		1019.63	2	2037.24
QDMAVDSAGKITYRIKLE	E	V	1.64	0.49	-0.909		1019.70	2	2037.39
QDMAVDSAGKITYRIKLE	E	V	2.75	0.448	3.06		1019.82	2	2037.63
QDMAVDSAGKITYRIKLE	E	V	2.73	0.415	3.7		1019.90	2	2037.79
QDMAVDSAGKITYRIKLE	E	V	1.95	0.422	0.444		1020.09	2	2038.16
QDMAVDSAGKITYRIKLE	E	V	2.52	0.333	0		1020.13	2	2038.24

*Key to fraction designation: *X* - Ammonium Sulfate cut A-F (see Materials and Methods), *Y* = AIEX fraction (13 sequential elution fractions, Roman Numeral I to XIII), "/" denotes pooled fractions, *Z* = RP-HPLC fraction (sequential elutions 1 through 1

Of considerable interest were the 15 fractions that demonstrated significant reactivity with only HIV+TB+ patients' sera. Western blot analysis of these fractions with HIV+TB+ patients' sera failed to demonstrate reactive bands. Furthermore, there were no protein bands in common between these fractions when analyzed by SDS-PAGE and silver staining, but treatment of these fractions with pronase (10 μg/ml for 60 min) prior to microarray printing significantly abrogated reactivity to patients' sera (data not shown). Together these observations suggest that 1) a single protein antigen may not be responsible for the reactivity of these 15 fractions, 2) a common antigen such as a small peptide may be responsible for the reactivity, or 3) the reactivity is due to a non-proteinaceous bacterial product complexed with protein.

3.4 Discussion

Previous studies from our laboratories employed 2-D immunoblotting to characterize the profile of *Mtb* proteins recognized by TB patients' sera (15, 16). However, this methodology is not well suited for analysis of large numbers of sera due to problems of reproducibility, difficulty in quantifying the results, and refractivity to high-throughput analyses. Protein microarrays offer a means by which a large number of sera can be analyzed not only to identify serologically reactive proteins, but to establish antigen recognition profiles based on the state or severity of disease (1). At present a complete recombinant protein library of *Mtb* does not exist. Therefore, to perform protein microarray studies for TB a novel approach involving a robust fractionation strategy that yielded 960 native protein fractions was utilized. The availability of such arrays allowed us to address differences in the patterns of antigens recognized by individuals exhibiting various forms of TB. Specifically, TB patients' sera recognized a

much greater number of protein fractions than did sera of PPD+ healthy controls and patients with noncavitary TB recognized only a subset (44%) of the fractions that reacted to patients with advanced, cavitary disease. This pattern of reactivity agrees well with our previous results obtained by 2-D Western blot analysis where only three to four CFPs reacted to sera of PPD+ healthy individuals and where 12 of the 26 cavitary TB reactive proteins (46%) were recognized by noncavitary TB patients (15, 16).

The inclusion of antigen identification into these current studies enabled a more in-depth assessment of the overlap between the three disease states (cavitary, noncavitary and HIV+TB+) studied. Of the 55 fractions recognized by both cavitary and noncavitary TB patients, 11 antigens (LAM, the 45 kDa Apa protein, the 19-kDa LpqH protein, Ag85A, Ag85B, Bfrb, GlcB, LppZ, Rv3881c, SecE2, and SodC) were identified as being serologically dominant. Of particular interest is the fact that all five fractions possessing GlcB as the reactive species were recognized by both noncavitary and cavitary TB patients, confirming previous reports that this antigen is recognized early in disease progression (16, 17). Our analyses also identified a total of 68 fractions recognized exclusively by cavitary TB patients' sera. Most of the antigens represented by these fractions (the 45-kDa Apa protein, Ag85B, LppZ, SecE2, SodC, LAM, Rv3881c, and the 19-kDa protein) overlapped with those recognized by both cavitary and noncavitary TB patients. However, four antigens (38-kDa PstS1 protein, HspX, MPT64, and TrxC) were recognized only by cavitary TB patients' sera, thus, providing several antigens that may be useful in demarcating cavitary and noncavitary TB patients. The identification of the 38-kDa PstS1 as a cavitary TB specific antigen concurs with previous reports that this antigen is recognized predominantly by patients with advanced disease (16, 17)

Previous experimental approaches have failed to identify an antigen that distinguishes noncavitary TB patients' from individual with other stages of the disease. In this study four native protein fractions were designated as noncavitary TB specific based on our selection parameters. A more in-depth inspection of these fractions, however, suggests they are likely not noncavitary TB specific because these four fractions were each recognized by four of eleven (36%) cavitary TB patients, just missing the 40% cutoff used to construct the Venn diagram. Additionally, the 45-kDa Apa protein and the 19-kDa protein were found to be the antigens responsible for the reactivity in three of these four fractions, and these same antigens were also found in five fractions recognized by sera of cavitary TB patients. Thus, a noncavitary TB specific antigen remains elusive and it appears that regardless of the methodology employed noncavitary TB patients react with a subset of those antigens recognized by cavitary TB patients.

While the overall data and conclusions obtained through these microarray-based studies were consistent with earlier work there were a few discrepancies with studies based on 2-D immunoblots (15, 16). In contrast to the qualitative 2-D immunoblot data the results obtained with microarrays were quantitative. Thus, the selection of reactive antigens or fractions was based on the percentage of sera with a NAvSNR value greater than an experimentally determined cutoff. This led to the exclusion of two antigens previously identified by 2-D immunoblots, but which fell outside the criteria set in this study to define significant reactivity. This was most notable with GlcB and its reactivity to HIV+TB+ patients' sera. GlcB was previously shown to react with HIV+TB+ sera in both 2-D immunoblot and ELISA formats (5, 15). However, when the microarray data

set was quantified the percentage of HIV+TB+ patients that recognized GlcB containing fractions was just below the 40% cutoff. Previous work shows that sera from this group of patients reacts to the same set of CFP as recognized by noncavitary TB patients (15). However, our present work defines fractions containing an unidentifiable antigen as the only material with significant reactivity to HIV+TB+ patients' sera. The exceptionally strong response of these fractions with HIV+TB+ sera likely led to a bias in data analysis for this patient group. A second difference with the 2-D Western blot data was the failure to define MPT51 as a dominant antigen. By 2-D PAGE MPT51 readily separates from LAM and is recognized as a dominant antigen in multiple disease states (15) . Probing the native protein array slide with the MPT51 specific MAb (IT52) demonstrated that MPT51 cofractionated with LAM (data not shown). Since LAM-containing fractions were excluded from further protein antigen analyses, MPT51 was not designated as a significant serological antigen. In our previous work there were approximately 12 protein spots that reacted to patients' sera by 2-D immunoblot, but were unidentifiable (15). Although it is possible that some of these 2-D protein spots correspond to the four novel antigens identified (see Chapter 4), it is not possible to draw such conclusions without further analyses.

Our earlier work on the serological response to *Mtb* proteins in human disease indicated considerable homogeneity in this response among TB patients (15). The previously accepted dogma of heterogeneity in antigen recognition likely resulted from the differences in the immunological response among the various states of the disease and poor immunological reactivity of *Mtb* recombinant proteins expressed in *E. coli* (9). The use of protein arrays has confirmed our previous observations and hypotheses.

Moreover, the ability to assess antigen recognition profiles between disease states allowed for the identification of several proteins recognized by both cavitary and noncavitary TB patients' sera and at least four proteins that appear diagnostic of cavitary TB.

3.5 Literature cited

1. **Davies, D. H., X. Liang, J. E. Hernandez, A. Randall, S. Hirst, Y. Mu, K. M. Romero, T. T. Nguyen, M. Kalantari-Dehaghi, S. Crotty, P. Baldi, L. P. Villarreal, and P. L. Felgner**. 2005. Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery. Proc Natl Acad Sci U S A **102**(3): 547-52.

2. **Dobos, K. M., K. H. Khoo, K. M. Swiderek, P. J. Brennan, and J. T. Belisle**. 1996. Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. J Bacteriol **178**(9): 2498-506.

3. **Hellman, U., C. Wernstedt, J. Gonez, and C. H. Heldin**. 1995. Improvement of an "In-Gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. Anal Biochem **224**(1): 451-5.

4. **Hirschfield, G. R., M. McNeil, and P. J. Brennan**. 1990. Peptidoglycan-associated polypeptides of *Mycobacterium tuberculosis*. J Bacteriol **172**(2): 1005-13.

5. **Laal, S., K. M. Samanich, M. G. Sonnenberg, J. T. Belisle, J. O'Leary, M. S. Simberkoff, and S. Zolla-Pazner**. 1997. Surrogate marker of preclinical tuberculosis in human immunodeficiency virus infection: antibodies to an 88-kDa secreted antigen of *Mycobacterium tuberculosis*. J Infect Dis **176**(1): 133-43.

6. **Laal, S., K. M. Samanich, M. G. Sonnenberg, S. Zolla-Pazner, J. M. Phadtare, and J. T. Belisle**. 1997. Human humoral responses to antigens of *Mycobacterium tuberculosis*: immunodominance of high-molecular-mass antigens. Clin Diagn Lab Immunol **4**(1): 49-56.

7. **Laal, S. and S. Y.A.**, Immune-Based Methods, in Tuberculosis and the Tubercle Bacillus, S.T. Cole, Editor. 2005, ASM Press: Washington, D.C. p. 71-83.

8. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**(5259): 680-5.

9. **Lyashchenko, K., R. Colangeli, M. Houde, H. Al Jahdali, D. Menzies, and M. L. Gennaro**. 1998. Heterogeneous antibody responses in tuberculosis. Infect Immun **66**(8): 3936-40.

10. **Mawuenyega, K. G., C. V. Forst, K. M. Dobos, J. T. Belisle, J. Chen, E. M. Bradbury, A. R. Bradbury, and X. Chen**. 2005. *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. Mol Biol Cell **16**(1): 396-404.

11. **McDonald, W. H. and J. R. Yates, 3rd**. 2002. Shotgun proteomics and biomarker discovery. Dis Markers **18**(2): 99-105.

12. **Morrissey, J. H.** 1981. Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. Anal Biochem **117**(2): 307-10.

13. **Sada, E., P. J. Brennan, T. Herrera, and M. Torres**. 1990. Evaluation of lipoarabinomannan for the serological diagnosis of tuberculosis. J Clin Microbiol **28**(12): 2587-90.

14. **Saeed, A. I., V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, and J. Quackenbush**. 2003. TM4: a free, open-source system for microarray data management and analysis. Biotechniques **34**(2): 374-8.

15. **Samanich, K., J. T. Belisle, and S. Laal**. 2001. Homogeneity of antibody responses in tuberculosis patients. Infect Immun **69**(7): 4600-9.

16. **Samanich, K. M., J. T. Belisle, M. G. Sonnenberg, M. A. Keen, S. Zolla-Pazner, and S. Laal**. 1998. Delineation of human antibody responses to culture filtrate antigens of *Mycobacterium tuberculosis*. J Infect Dis **178**(5): 1534-8.

17. **Samanich, K. M., M. A. Keen, V. D. Vissa, J. D. Harder, J. S. Spencer, J. T. Belisle, S. Zolla-Pazner, and S. Laal**. 2000. Serodiagnostic potential of culture filtrate antigens of *Mycobacterium tuberculosis*. Clin Diagn Lab Immunol **7**(4): 662-8.

18. **Singh, K. K., Y. Dong, J. T. Belisle, J. Harder, V. K. Arora, and S. Laal**. 2005. Antigens of *Mycobacterium tuberculosis* recognized by antibodies during incipient, subclinical tuberculosis. Clin Diagn Lab Immunol **12**(2): 354-8.

19. **Singh, K. K., Y. Dong, L. Hinds, M. A. Keen, J. T. Belisle, S. Zolla-Pazner, J. M. Achkar, A. J. Nadas, V. K. Arora, and S. Laal**. 2003. Combined use of serum and urinary antibody for diagnosis of tuberculosis. J Infect Dis **188**(3): 371-7.

20. **Singh, K. K., Y. Dong, S. A. Patibandla, D. N. McMurray, V. K. Arora, and S. Laal**. 2005. Immunogenicity of the *Mycobacterium tuberculosis* PPE55 (Rv3347c) protein during incipient and clinical tuberculosis. Infect Immun **73**(8): 5004-14.

21. **Smith, P. K., R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson, and D. C. Klenk**. 1985. Measurement of protein using bicinchoninic acid. Anal Biochem **150**(1): 76-85.

22. **Sonnenberg, M. G. and J. T. Belisle**. 1997. Definition of *Mycobacterium tuberculosis* culture filtrate proteins by two-dimensional polyacrylamide gel electrophoresis, N-terminal amino acid sequencing, and electrospray mass spectrometry. Infect Immun **65**(11): 4515-24.

Chapter IV

Discovery and Validation of Tuberculosis-Serodiagnostic Biomarkers

Partially presented in Mark J. Sartain, Richard A. Slayden, Krishna K. Singh, Suman Laal, and John T. Belisle. 2006. Disease State Differentiation and Identification of Tuberculosis Biomarker via Native Antigen Array Profiling. Molecular and Cellular Proteomics. Nov; Vol 5 (11): 2102-2113

4.1 Introduction

A critical element of TB control is the early and sensitive diagnosis of infection and disease. Development of immune-based tests that are based on the humoral (serological) antibody response against *Mtb* may be the most realistic means to provide developing countries with a rapid, low-cost, and simply executed tool for TB diagnosis. However, the scientific community has slowly come to the realization that an effective assay must be based on multiple antigens. Therefore, recent efforts have focused on identifying and evaluating the most promising serodiagnostic antigens and combinations thereof. Through the work of multiple laboratories greater than 16 *Mtb* proteins have been identified as potential serodiagnostic antigens, each producing varying degrees of diagnostic sensitivity (17). Many of these antigens were identified as a direct result of their *in vitro* abundance and ease of biochemical purification; therefore, new strategies are needed to continue to identify novel antigens with serodiagnostic utility. Chapter 3 described previously known *Mtb* antigens identified with a combination of a robust protein separations strategy and emerging protein microarray technology. This methodology also identified four novel *Mtb* antigens with serodiagnostic utility, and the identification, production, characterization, and validation of these antigens is the focus of this chapter.

4.2 Materials and Methods

4.2.1 Bacterial growth and subcellular fractionation

E. coli strain TOP10 (Invitrogen) containing various recombinant plasmids was grown on Luria-Bertani (LB) agar containing kanamycin (25μg ml^{-1}). Growth of *E. coli* strain BL21(DE3) for recombinant protein purification was performed in LB broth containing 1% glucose and 100 μg/ml ampicillin.

Recombinant clones of *M. smegmatis* mc^2155 were selected on LB agar containing kanamycin (25 μg ml^{-1}). Growth of *M. smegmatis* for protein purification was achieved by propagation in 12 L of GAS medium (35) containing kanamycin (25 ug ml^{-1}) at 37°C with gentle shaking. Cells were harvested at three days of growth, and bacterial pellets were suspended in PBS and passed through a French press four times at 1500 psi to generate WCL.

Growth of *Mtb* strain H37Rv for isolation of subcellular fractions was achieved by propagation in 14 L of GAS medium at 37°C with gentle shaking. Cells were harvested at 14 days of growth, and individual subcellular fractions of cytosol, membrane, cell wall, and CF were isolated as previously described (14, 32). Final protein concentrations were determined using the BCA protein assay (30).

4.2.2 Construction of recombinant plasmids

Recombinant plasmid constructs were created according to standard protocols (27). PCR amplifications were performed with *PfuTurbo* DNA polymerase (Stratagene) using *Mtb* H37Rv genomic DNA as the template. All PCR products were first cloned into pCR4Blunt-TOPO according to the manufacturer's protocols (Invitrogen), and subsequently recovered by restriction enzyme digestions that targeted the restriction site

linker sequences designed into the primers (Table 4.1). To generate *E. coli* and mycobacterial expression constructs for recombinant production of SecE2 (rSecE2), the *rv0379* gene was amplified using primer pair SecE2F and SecE2R. The 213 bp fragment isolated from the pCR4Blunt intermediate plasmid construct by digestion with *Nde*I/*Hind*III was ligated into the *Nde*I/*Hind*III sites of pET23b and pVV16 to generate pMRLB56 and pMRLB57, respectively. To generate *E. coli* and mycobacterial expression constructs for recombinant production of TrxC (rTrxC), the *rv3914* gene was amplified using primer pair TrxCF and TrxCR. The 348 bp fragment isolated from the pCR4Blunt intermediate plasmid construct by digestion with *Nde*I/*Hind*III was ligated into the *Nde*I/*Hind*III sites of pET23b and pVV16 to generate pMRLB58 and pMRLB59, respectively. To generate the *E. coli* expression construct for recombinant production of mature LppZ lacking a signal peptide (SP-rLppZ), the *rv3006* gene minus the region encoding the N-terminal signal peptide was amplified using primer pair LppZ(-SP)F and LppZR1. The 1053 bp fragment isolated from the pCR4Blunt intermediate plasmid construct by digestion with *Nde*I/*Xho*I was ligated into the *Nde*I/*Xho*I sites of pET23b to generate pMRLB54. To generate a mycobacterial-expression construct that encoded mature and fully modified recombinant LppZ (rLppZ), the full-length *rv3006* gene was amplified using primer pair LppZ(+SP)F and LppZR2. The 1119 bp fragment isolated from the pCR4Blunt intermediate plasmid construct by digestion with *Nde*I/*Hind*III was ligated into the *Nde*I/*Hind*III sites of pVV16 to generate pMRLB55. To generate the *E. coli* expression construct for recombinant production of mature SodC lacking a signal peptide (SP-rSodC), the *rv0432* gene minus the region encoding the N-terminal signal peptide was amplified using primer pair SodC(-SP)F and SodCR1. The 627 bp fragment

isolated from the pCR4Blunt intermediate plasmid construct by digestion with *Nde*I/*Xho*I was ligated into the *Nde*I/*Xho*I sites of pET23b to generate pMRLB60. To generate a mycobacterial-expression construct that encoded mature and fully modified recombinant SodC (rSodC), the full-length *rv0432* gene was amplified using primer pair SodC(+SP)F and SodCR2. The 720 bp fragment isolated from the pCR4Blunt intermediate plasmid construct by digestion with *Nde*I/*Hind*III was ligated into the *Nde*I/*Hind*III sites of pVV16 to generate pMRLB61. The pMRLB5 construct, designed for recombinant production of BfrB (rBfrB) in *E. coli*, was obtained from the Colorado State University TB Research Materials and Vaccine Testing (CSU TBVTRM) Contract, NIAID, National Institutes of Healthy Contract NO1 AI-75320. To generate a mycobacterial-expression construct for recombinant production of BfrB, pMRLB5 was digested with *Nde*I/*Hind*III and the recovered 543 bp fragment was ligated into the *Nde*I/*Hind*III sites of pVV16 to generate pMRLB63. All plasmid constructs were confirmed by nucleotide sequencing through Macromolecular Resources (Fort Collins, CO).

Table 4.1 Plasmids and primers used in this study

	Description/Sequence	Reference or Source
Plasmid		
pCR4Blunt-TOPO	cloning vector, Amp^r, Kan^r	Invitrogen
pET23b	*E. coli* expression vector, Amp^r	Novagen
pVV16	*Mycobacteria* expression vector, Kan^r, Hyg^r	Schulbach *et al.* (2001)
pMRLB5	*bfrb (rv3841)* gene in pet23b	CSU TBVTRM Contract
pMRLB54	*lppZ (rv3006)* gene fragment lacking signal peptide sequence in pet23b	This study
pMRLB55	full-length *lppZ (rv3006)* gene in pVV16	This study
pMRLB56	*secE2 (rv0379)* gene in pet23b	This study
pMRLB57	*secE2 (rv0379)* gene in pVV16	This study
pMRLB58	*trxC (rv3914)* gene in pet23b	This study
pMRLB59	*trxC (rv3914)* gene in pVV16	This study
pMRLB60	*sodC (rv0432)* gene fragment lacking signal peptide sequence in pet23b	This study
pMRLB61	full-length *sodC (rv0432)* gene in pVV16	This study
pMRLB63	*bfrb (rv3841)* gene in pVV16	This study
Primer		
LppZ(-SP)F	5'-CATATGTGCGCACGGTTCAAC	This study
LppZR1	5'-CTCGAGGGTCTTGTCGTCGTT	This study
LppZ(+SP)F	5'-CATATGTGGACAACGCGGTTG	This study
LppZR2	5'-AAGCTTGGTCTTGTCGTCGTT	This study
SecE2F	5'-CATATGAGTGTGTACAAGGTGATCG	This study
SecE2R	5'-AAGCTTGCGCGGTTGCGCC	This study
TrxCF	5'-CATATGACCGATTCCGAGAAG	This study
TrxCR	5'-AAGCTTGTTGAGGTTGGGAAC	This study
SodC(-SP)F	5'-CATATGTGCTCGTCGCCGCAG	This study
SodCR1	5'-CTCGAGGCCGGAACCAATGAC	This study
SodC(+SP)F	5'-CATATGCCAAAGCCCGCCGAT	This study
SodCR2	5'-AAGCTTGCCGGAACCAATGAC	This study

Restriction endonuclease sites are underlined

4.2.3 Recombinant protein purification

For production of recombinant proteins in *E. coli*, plasmids pMRLB5, pMRLB54, pMRLB56, pMRLB58, and pMRLB60 were transformed into *E. coli* BL21(DE3) (33). Two-liter cultures were grown at 37°C to an OD_{600} of 0.3-0.6, and the recombinant genes were expressed via addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 4-6 h. The cells were harvested, lysed with lysozyme and probe sonication in the presence of RNase, DNAse, and protease inhibitors, and centrifuged at 16,000 X *g* for 30

min. The rBfrB and rTrxC supernatants were equilibrated in denaturing binding buffer (5.0 mM imidazole, 0.5 M $NaCl_2$, 6.0 M urea, 20 mM Tris-Cl pH 7.9), while the rSecE2, SP-rSodC, and SP-rLppZ proteins formed inclusion bodies and were solubilized with denaturing binding buffer. The supernatants and the solubilized protein samples were applied to a 1 mL column packed with His-bind resin (Novagen) equilibrated in the denaturing binding buffer. The recombinant proteins were purified by washing with 20 column volumes (CV) of denaturing binding buffer, 25 CV of denaturing wash buffer (60 mM imidazole, 0.5 M $NaCl_2$, 6.0 M urea, 20 mM Tris-Cl pH 7.9), 10 CV of 10 mM Tris buffer (pH 8.0), 10 CV of 0.5% ASB-14 (Calbiochem) in 10 mM Tris buffer (pH 8.0), and again in 10 CV of 10 mM Tris buffer (pH 8.0) to remove detergent. Proteins were eluted from the column by the addition of 10 CV of denaturing elution buffer (1.0 M imidazole, 0.5 M $NaCl_2$, 6.0 M urea, 20 mM Tris-Cl pH 7.9).

For production of recombinant proteins in *M. smegmatis*, the plasmids pMRLB57, pMRLB59, pMRLB61, and pMRLB63 were electroporated into *M. smegmatis* mc^2155 by the method of Snapper, *et al.* (31). Cells from 72-hour (mid- to late-log) cultures of recombinant *M. smegmatis* were washed three times with PBS (pH 7.4), resuspended in denaturing binding buffer containing protease inhibitors, and passed through a French press five times at 1500 p.s.i. The resulting lysates were centrifuged at 27,000 X *g* for 30 min and the supernatants were applied to a 1 mL column packed with His-bind resin equilibrated in denaturing binding buffer. The recombinant proteins were purified by washing the column with 20 CV of denaturing binding buffer, 25 CV of denaturing wash buffer (60 mM imidazole, 0.5 M $NaCl_2$, 6.0 M urea, 20 mM Tris-Cl pH 7.9), 10 CV of 10 mM Tris buffer (pH 8.0), 10 CV of 0.5% ASB-14 (Calbiochem) in 10 mM Tris buffer

(pH 8.0), and again in 10 CV of 10 mM Tris buffer (pH 8.0) to remove detergent. Proteins were eluted from the column by the addition of 10 CV of denaturing elution buffer (1.0 M imidazole, 0.5 M $NaCl_2$, 6.0 M urea, 20 mM Tris-Cl pH 7.9).

Each fraction was extensively dialyzed against 10 mM ammonium bicarbonate and concentrated ten-fold by ultrafiltration. Precipitates of rSecE2 from *E. coli* and *M. smegmatis* were solubilized by heating at 95°C in the presence of 1% SDS. All protein concentrations were determined using the BCA protein assay (30). MS/MS-based protein identifications were conducted as described in section 3.2.7.

4.2.4 Human sera and antibodies

Serum samples obtained with informed consent from 93 individuals were included in these studies. These individuals can be categorized into the following groups.

(i) *Fifteen PPD+ healthy individuals.* Twelve of these individuals were recent immigrants from areas where TB is endemic, many of whom had been vaccinated with *M. bovis* BCG, and it is likely that a high proportion would also have been exposed to *M. tuberculosis*. The remaining three PPD+ healthy individuals were from the United States and were not BCG-vaccinated.

(ii) *Five PPD-negative (PPD-) individuals.* These were healthy individuals working at the MVAMC. These individuals were unlikely to be infected with *Mtb*.

(iii) *Thirty-seven cavitary TB patients, with moderate to advanced cavitary lesions as determined by chest X-rays.* These were AFB sputum-smear positive patients obtained from the LRSITRD who were all bled prior to initiation of therapy for TB. None of these patients were HIV infected.

(iv) *Twenty HIV+TB+ patients.* These were all AFB sputum smear-positive patients from the Post Graduate Institute of Medical Education and Research, Chandigarh, India. Ten of these patients had normal chest X-rays, two had cavitary lesions, four had miliary, three had infiltration and one had interstitial infiltration with PCP. Among these patients ten had pulmonary TB and the remaining ten had extra-pulmonary TB. These patients were bled either prior to or within two weeks of the initiation of therapy for TB.

(v) *Sixteen household contacts of smear-positive TB patients.* These were clinically asymptomatic household contacts of infectious, untreated, smear-positive TB patients, and these sera were also obtained from the LRSITRD. Two of these individuals were PPD+, three were PPD-, and the PPD statuses of the remaining 11 were not known. However, despite being asymptomatic, because of their frequent exposure to smear-positive family members a small proportion of these individuals (estimated to be from 6 to 29% in different studies) are expected to have an active infection (5, 10, 11, 36).

Rabbit polyclonal sera to SecE2, TrxC, BfrB, SodC, and LppZ were generated by Strategic Biosolutions (Ramon, CA) in a standard rabbit protocol. Purified SecE2, BfrB, SP-rSodC, and SP-rLppZ from *E. coli* were used as the antigens. The polyclonal antisera against LAM and mAb CS-35 were obtained from the CSU TBVTRM contract.

4.2.5 SDS-PAGE and Western blot analyses

Purified LAM from *Mtb* was obtained from the CSU TBVTRM Contract. Aliquots of protein or LAM were subjected to SDS-PAGE using 10-20% Tricine gels (Invitrogen). Gels were stained with Coomassie Brilliant Blue R250 (7) or silver nitrate (20), or electroblotted to nitrocellulose membranes (Bio-Rad) as previously described

(32). The nitrocellulose membranes were blocked with 3% nonfat milk in PBS (pH 7.2) for 2 h and exposed overnight to anti-His_5 monoclonal antibody (Qiagen, 1:1000), SodC antiserum (1:2000), LppZ antiserum (1:5000), BfrB antiserum (1:50,000), or SecE2 antiserum (1:1000) each diluted in 1% nonfat milk in PBS (pH 7.2). The blots were washed with PBS (pH 7.2) containing 0.1% Tween 20, probed for 1.5 h with alkaline phosphatase-conjugated anti-mouse IgG whole molecule (Sigma) or anti-rabbit IgG Fc fragment (Calbiochem) both diluted 1:2000 in 1% nonfat milk in PBS (pH 7.2), and washed extensively. Antigen-antibody complexes were visualized by color development with 5-bromo-4-chloro-indoyl-phosphatase-nitroblue tetrazolium substrate (Kirkegaard & Perry Laboratories, Gaithersburg, MD). Western blot analysis of the multidimensional chromatography fractions with pooled patients' sera is described in section 3.2.6.

4.2.6 ELISA

LAM-free CFP (LFCFP) and rGlcB/Rv1837c from *E. coli* (29) were obtained from the CSU TBVTRM Contract. For assessment of the reactivity of antigen preparations with sera from TB patients and control individuals, optimal plate coating concentrations were determined for each antigen. LFCFP was coated at 5 μg/ml and each recombinant protein was coated at 4 μg/ml (50 μg/well) overnight at 4°C in coating buffer (3.1 mM NaN_3, 15 mM Na_2CO_3, 35 mM $NaHCO_3$, pH 9.6). The plates (Immulon 2HB, Dynex) were washed three times with PBS (pH 7.1) containing 0.1% Tween 20, and the wells were blocked with 1% BSA-PBS containing 0.1% Tween 20 (blocking buffer) for 2 h at 37°C. After four washes, the antigen-coated wells were exposed to sera (1:50 in 0.1X blocking buffer) for 30 min at 37°C, and washed six times with PBS (pH 7.1) containing 0.1% Tween 20. The wells were exposed to a mixture of diluted alkaline

phosphatase-conjugated protein A (1:2,000; Sigma) and goat anti-human IgA (1:1,000; Sigma) for 30 min at 37°C. The wells were washed eight times with Tris-buffered saline (150 mM NaCl, 50 mM Tris, pH 7.4), and the color was developed using an Invitrogen amplification system (Invitrogen). The mean optical density at 490 nm (OD_{490}) obtained with sera from healthy individuals (combined PPD- and PPD+ controls) plus three SD was used as the cutoff to determine positive responses in the patients.

4.3 Results

4.3.1 Discovery of novel antigens with potential serodiagnostic roles

The MS/MS-based approach to antigen identification described in chapter three led to the elucidation of four new *Mtb* B cell antigens (Table 4.2). The first of these, SodC (Rv0432), a 27-kDa Cu,Zn superoxide dismutase (38), was the sole reactive constituent of four fractions originating from the CFP pool (B-III-3, B-III-4, B-IV-3, and B-IV-4), and was significantly recognized by both noncavitary- (five out of nine) and cavitary-TB (eight out of eleven) sera (Fig. 1A). MS/MS analysis of the corresponding protein in fraction B-III-4 resulted in 42% amino-acid coverage of the predicted protein sequence encoded by ORF *rv0432* (Table 4.3).

LppZ (Rv3006) was identified as the only reactive product of five fractions (A-III-5, A-IV-4, A-IV-5, A-IV-6, and A-V-5) originating from the CFP pool, and was strongly recognized by noncavitary- (four out of nine) and cavitary-TB (six out of eleven) sera. These fractions were generated under similar separation conditions, and Western blot analysis showed all possessed a dominant reactive 45-kDa product (Fig. 1B). MS/MS analysis of the corresponding protein in fraction A-IV-5 resulted in 37% amino-acid coverage of the predicted protein sequence encoded by ORF *rv3006* (Table 4.3).

Bfrb (Rv3841), a putative iron-storage protein (6), and a previously described T cell antigen (8), was identified as the reactive protein in two fractions (A-VII-1 and A-VIII-1) with significant reactivity to four out of nine noncavitary- and six out of eleven cavitary-TB sera. MS/MS analysis of the 20-kDa reactive product (Fig. 1C) in fraction A-VIII-1 resulted in 37% amino-acid coverage of the predicted protein sequence encoded by ORF *rv3841* (Table 4.3), and analysis of the silver-stained polyacrylamide gels revealed the 20-kDa band was the sole protein constituent in both fractions (data not shown).

TrxC (Rv3914), a 12-kDa thioredoxin (37), was found along with the 45-kDa Apa antigen to account for the serological reactivity of fraction B-II-9. In contrast to the other novel-antigen-containing fractions, fraction B-II-9 was only significantly recognized by cavitary-TB patients' sera (six out of eleven). MS/MS analysis of the 12-kDa protein (Fig. 1D) resulted in 29% amino-acid coverage of the predicted protein sequence encoded by ORF *rv3914* (Table 4.3). Whether or not the significant reactivity of fraction B-II-9 was due to TrxC or the strongly seroreactive 45-kDa antigen (26) could not be determined.

When serological data were combined for those fractions containing the four novel antigens it was found that 56 and 91% of the noncavitary and cavitary TB patients' sera, respectively, showed positive reactivity. In comparison, when the data were combined for all fractions containing novel and previously identified protein antigens (see chapter 3), 78% of noncavitary and 91% of cavitary patients displayed reactivity.

Table 4.2 Novel antigens composing reactive fractions.

Antigen(s)	Fractionation Conditions (X,Y,Z)[a]	Patient Reactivity[b] HIV+TB+	Noncavitary TB	Cavitary TB	MS/MS[c]
Bfrb/Rv3841	A-VII-1	3/10 (30%)	4/9 (44%)	6/11 (55%)	
	A-VIII-1	3/10 (30%)	4/9 (44%)	6/11 (55%)	(37)
LppZ/Rv3006	A-III-5	2/10 (20%)	3/9 (33%)	6/11 (55%)	
	A-IV-4	2/10 (20%)	4/9 (44%)	5/11 (44%)	
	A-IV-5	2/10 (20%)	3/9 (33%)	6/11 (55%)	(37)
	A-IV-6	2/10 (20%)	3/9 (33%)	6/11 (55%)	
	A-V-5	2/10 (20%)	4/9 (44%)	5/11 (44%)	
SodC/Rv0432	B-III-3	2/10 (20%)	4/9 (44%)	6/11 (55%)	
	B-III-4	2/10 (20%)	5/9 (56%)	8/11 (73%)	(42)
	B-IV-3	2/10 (20%)	4/9 (44%)	7/11 (64%)	
	B-IV-4	3/10 (30%)	3/9 (33%)	7/11 (64%)	
TrxC/Rv3914 & Apa/Rv1860[d]	B-II-9	2/10 (20%)	3/9 (33%)	6/11 (55%)	(29)[e]

[a] Key to fraction designation: *X* = Ammonium Sulfate cut A-F (see Materials and Methods), *Y* = AIEX fraction (13 sequential elution fractions, Roman Numeral I to XIII), "/" denotes pooled fractions, *Z* = RP-HPLC fraction (sequential elutions 1 through 15).

[b] Number (Percentage) of patient sera significantly reactive against fraction in question. Significant is > 3 X SD above PPD+ mean.

[c] Antigen identification based on MS/MS analysis. Percent amino acid sequence coverage is in parentheses.

[d] The Apa was identified with the MAb CS-93.

[e] The percent amino acid coverage is for TrxC

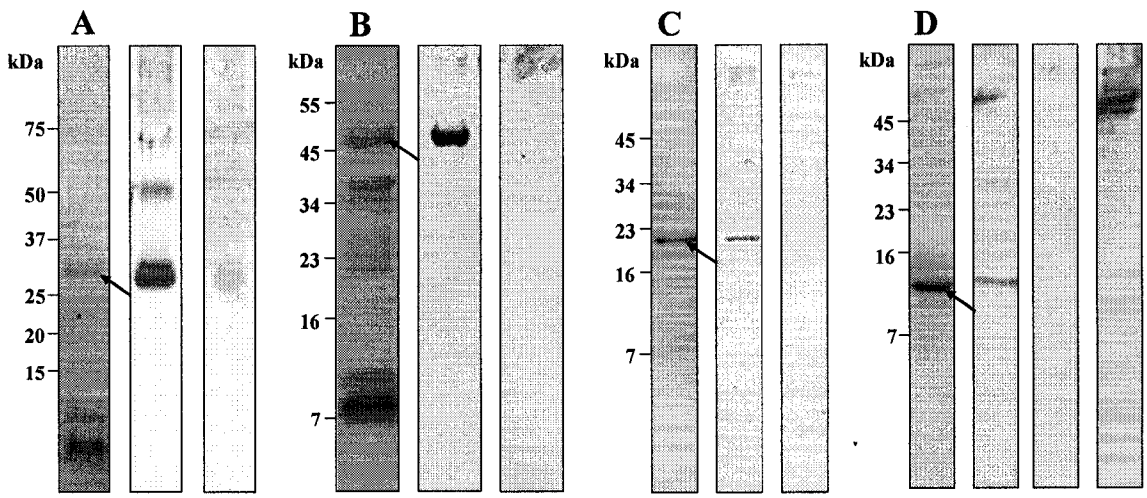


Fig. 4.1 Identification of novel serodiagnostic antigens. Significantly reactive fractions identified by microarray analyses were analyzed by SDS-PAGE and coomassie staining (panel 1), and Western blot with pooled sera from cavitary TB (panel 2) and healthy PPD+ individuals (panel 3). The coomassie-stained proteins (marked with ←) corresponding to reactive bands were digested with trypsin and the antigens were identified with MS/MS-based methods (see Table 4.3). (**A**) The single reactive 27-kDa product of fraction B-III-4 was identified as SodC/Rv0432. (**B**) The single reactive 45-kDa product of fraction A-IV-5 was identified as LppZ/Rv3006. (**C**) The single reactive 20-kDa product of fraction A-VIII-1 was identified as BfrB/Rv3841. (**D**) A reactive 12-kDa product of fraction B-II-9 was identified as TrxC/Rv3914, and a second reactive 50-kDa product was identified as the 45-kDa/Apa antigen using α-45 kDa polyclonal sera (panel 4).

Table 4.3 MS-based identification of novel antigens

Protein accession number: Rv0432 Molecular Weight (Da): 23826.2
Common Name(s): SodC AA Coverage: 42% # of total spectra: 15 # of unique spectra: 6 # of unique peptides: 6
Fraction: B-III-4

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
LTPGFHGLHIHQVGK	K	C	2.09	0.433	0		821.17	2	1640.33
YVQVNGTPGPDETTLTTGDAGK	R	R	4.26	0.623	7.26		1111.48	2	2220.95
YVQVNGTPGPDETTLTTGDAGK	R	R	0	0	4.32		1113.23	2	2224.44
TAIIIHAGADNFANIPPER	K	Y	3.75	0.68	5.62		1010.54	2	2019.06
TAIIIHAGADNFANIPPER	K	Y	4.12	0.673	12.2		1010.68	2	2019.34
TAIIIHAGADNFANIPPER	K	Y	3.76	0.666	5.96		1010.81	2	2019.6
TAIIIHAGADNFANIPPER	K	Y	4.25	0.637	10.8		1011.19	2	2020.37
TAIIIHAGADNFANIPPER	K	Y	2.38	0.593	3.59		1011.20	2	2020.38
YVQVNGTPGPDETTLTTGDAGKR	R	V	3.06	0.525	2.55		1190.19	2	2378.37
GDGSAMLVTTTDAFTMDDLLSGAK	R	T	4.25	0.699	0		1209.11	2	2416.2
GDGSAMLVTTTDAFTMDDLLSGAK	R	T	4.28	0.693	0	Oxidation (+16)	1217.13	2	2432.25
GDGSAMLVTTTDAFTMDDLLSGAK	R	T	0	0	2.39	2 X Oxidation (+16)	1225.04	2	2448.07
GDGSAMLVTTTDAFTMDDLLSGAK	R	T	4.53	0.634	4.72	2 X Oxidation (+16)	1225.11	2	2448.21
GDGSAMLVTTTDAFTMDDLLSGAK	R	T	0	0	4.35	2 X Oxidation (+16)	1226.65	2	2451.28
FEFANGYATVTIATTGVGK	K	L	2.45	0.555	2.57		974.54	2	1947.07

Protein accession number: Rv3006 Molecular Weight (Da): 38734.2
Common Name(s): LppZ AA Coverage: 37% # of total spectra: 30 # of unique spectra: 12 # of unique peptides: 10
Fraction: A-IV-5

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
KDTHAHAWALR	R	M	3.01	0.545	5.28		653.56	2	1305.1
VADGDIPKDILTGIPK	R	G	2.92	0.588	0		1653.68	1	1652.67
VADGDIPKDILTGIPK	R	G	3.29	0.609	2.85		826.77	2	1651.53
LAPSTGAVTGEPDVVR	R	K	3.36	0.616	4.22		785.00	2	1567.98
VKTVIPVDPAGDGGLMDIVLSPTYSQDR	K	L	2.96	0.538	0		1472.58	2	2943.15
VKTVIPVDPAGDGGLMDIVLSPTYSQDR	K	L	2.61	0.46	5.14	Oxidation (+16)	1480.67	2	2959.32
VKTVIPVDPAGDGGLMDIVLSPTYSQDR	K	L	3.53	0.365	1.89	Oxidation (+16)	987.41	3	2959.22
LMYAYISTPTDNR	R	V	3.63	0.583	4.21	Oxidation (+16)	781.10	2	1560.19
LMYAYISTPTDNR	R	V	3.26	0.647	5.8	Oxidation (+16)	781.54	2	1561.07
TVIPVDPAGDGGLMDIVLSPTYSQDR	K	L	2.61	0.239	2.09	Oxidation (+16)	1367.03	2	2732.05
TVIPVDPAGDGGLMDIVLSPTYSQDR	K	L	1.69	0.379	3.48	Oxidation (+16)	1367.12	2	2732.23
TAGDAEKLDDVVFPLFPQGGGFPR	K	N	4.79	0.603	5.43		1267.25	2	2532.49
TAGDAEKLDDVVFPLFPQGGGFPR	K	N	4.8	0.501	4.1		1267.46	2	2532.9
TAGDAEKLDDVVFPLFPQGGGFPR	K	N	3.01	0.576	2.59		1267.54	2	2533.06
TAGDAEKLDDVVFPLFPQGGGFPR	K	N	4.3	0.607	6.5		1267.67	2	2533.33
TAGDAEKLDDVVFPLFPQGGGFPR	K	N	2.57	0.363	2.39		1267.90	2	2533.79
ITGAVEEISISAEPK	R	V	3.38	0.539	5.2		772.44	2	1542.86
ITGAVEEISISAEPK	R	V	3.81	0.518	7.62		772.48	2	1542.95
ITGAVEEISISAEPK	R	V	3.03	0.55	7.92		772.43	2	1542.85
ITGAVEEISISAEPK	R	V	4.18	0.562	7.54		772.55	2	1543.08
ITGAVEEISISAEPK	R	V	3.4	0.508	4.96		772.55	2	1543.08
ITGAVEEISISAEPK	R	V	4.41	0.548	7.92		772.52	2	1543.02
ITGAVEEISISAEPK	R	V	3.4	0.53	6.24		772.59	2	1543.17
ITGAVEEISISAEPK	R	V	3.72	0.184	3.36		772.69	2	1543.36
ITGAVEEISISAEPK	R	V	3.41	0.467	5.14		772.73	2	1543.44
LAPSTGAVTGEPDVVRK	R	D	3.85	0.662	2.37		849.24	2	1696.47
MSPDGNVWGATVNK	R	T	2.87	0.416	0		738.21	2	1474.41
MSPDGNVWGATVNK	R	T	3.01	0.696	0		738.53	2	1475.04
MSPDGNVWGATVNK	R	T	3	0.533	4.96	Oxidation (+16)	746.23	2	1490.44
MSPDGNVWGATVNK	R	T	3.06	0.667	5.09	Oxidation (+16)	746.96	2	1491.9

Protein accession number: Rv3841 Molecular Weight (Da): 20423.6
Common Name(s): BfrB AA Coverage: 37% # of total spectra: 11 # of unique spectra: 7 # of unique peptides: 6
Fraction: A-VIII-1

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
EVDVAPAASGAPHAAGGR	R	-	2.87	0.528	8.08		816.87	2	1631.72
EVDVAPAASGAPHAAGGRL	R	-	2.94	0.571	5.17		873.74	2	1745.46
AGANLFELENFVAR	R	E	2.41	0.477	0		1552.48	1	1551.47
AGANLFELENFVAR	R	E	2.35	0.452	0		1552.67	1	1551.66
AGANLFELENFVAR	R	E	4.02	0.686	5.55		776.24	2	1550.46
NHAMMLVQHLLDR	R	D	2.3	0.528	0	Oxidation (+16)	797.88	2	1593.74
NHAMMLVQHLLDR	R	D	2.62	0.531	4.54	2 X Oxidation (+16)	805.45	2	1608.88
NHAMMLVQHLLDR	R	D	2.98	0.524	2.57	2 X Oxidation (+16)	805.55	2	1609.09
EALALALDQER	R	T	2	0.343	2.96		1228.57	1	1227.56
EALALALDQER	R	T	2.29	0.462	0.824		1230.51	1	1229.5
HFYSQAVEER	K	N	2.54	0.425	5.05		634.00	2	1265.98

Protein accession number: Rv3914 Molecular Weight (Da): 12526.5
Common Name(s): TrxC AA Coverage: 29% # of total spectra: 4 # of unique spectra: 3 # of unique peptides: 3
Fraction: B-II-9

Peptide sequence	Previous AA	Next AA	SEQUEST XCorr score	SEQUEST DCn score	X! Tandem -log(e) score	Modifications identified by spectrum	Precursor Mass	Spectrum charge	Actual peptide mass (AMU)
NFQVVSIPTLILFK	R	D	1.97	0.447	5.08		810.04	2	1618.07
NFQVVSIPTLILFK	R	D	1.92	0.0839	2.85		810.04	2	1618.07
NFQVVSIPTLILFKDGQPVKR	R	I	3.27	0.622	6.64		1200.33	2	2398.65
MVAPVLEEIATER	K	A	3.24	0.637	0		729.54	2	1457.07

*Key to fraction designation: *X* - Ammonium Sulfate cut A-F (see Materials and Methods), *Y* = AIEX fraction (13 sequential elution fractions, Roman Numeral I to XIII), "/" denotes pooled fractions, *Z* = RP-HPLC fraction (sequential elutions 1 through 15)

4.3.2 Production and initial characterization of novel antigens

In order to validate the seroreactivity of the newly described antigens identified with the microarray approach, recombinant forms of BfrB, LppZ, SodC, and TrxC were produced. Each recombinant protein was produced in *E. coli* and *M. smegmatis* in order to evaluate the effects of differing recombinant protein production systems on antibody recognition. Two of the four novel antigens identified, SodC/Rv0432 and LppZ/Rv3006, are putative lipoproteins with predicted type II signal peptide sequences (9, 34). The ability of *E. coli* to consistently recognize and correctly process mycobacterial signal sequences is unknown; therefore, the *lppZ* and *sodC* genes were cloned with and without their native signal sequences for protein production in *M. smegmatis* and *E. coli,* respectively. In addition to the four novel antigens identified in section 4.3.1, recombinant SecE2/Rv0379/Mtb8 was produced for serological evaluation in an attempt to evaluate a recent report that this antigen makes a good complementary component in multi-antigen serodiagnostic tests (15). To facilitate purification all recombinant proteins were fused in-frame with a C-terminal hexa-His tag.

The expression of the recombinant genes was assessed in *E. coli* and *M. smegmatis*. The previously known antigen rSecE2 and each of the four novel antigens (rBfrB, rLppZ, rSodC, and rTrxC) were detected in the *E. coli* cellular lysates by Western blot using an anti-His_5 antibody, while four of the five proteins (rSecE2, rTrxC, rBfrB, and rSodC) were detected in cellular lysates of *M. smegmatis*. Overexpression of rLppZ appeared toxic to *M. smegmatis* cultures, resulting in noticeable cellular lysis and a slowed growth rate compared to a vector control. Furthermore, only minute amounts of rLppZ protein were detected and purification attempts were unsuccessful. The

solubilities of the five recombinant proteins in *E. coli* were assessed, and SP-rLppZ and rSecE2 were found to be located in the insoluble pellet following cell lysis and centrifugation. These proteins were likely located in inclusion bodies; therefore, urea was employed for their solubilization and purification. All of the recombinant proteins produced in *E. coli* and *M. smegmatis* were purified in the presence of urea to keep downstream seroreactivity comparisons consistent.

The purified recombinant proteins were analyzed by SDS-PAGE and Western blot using the anti-His_5 antibody. A dominant ~11-kDa anti-His_5-reactive product was observed for rSecE2 from both *E. coli* and *M. smegmatis*, and was close in size to the expected molecular weights (MWs) of 9.5 and 9.0 kDa, respectively (Fig. 4.2A) A number of additional anti-His_5-reactive products were observed, including a dominant ~250-kDa band for rSecE2 from *E. coli*. The nature of these additional products is unknown, but may represent aggregates of rSecE2, consistent with the need to employ SDS to solubilize precipitates of rSecE2. A dominant ~14-kDa anti-His_5-reactive product was observed for rTrxC from both *E. coli* and *M. smegmatis*, and was close in size to the expected MWs of 14.1 and 13.6 kDa, respectively (Fig. 4.2B). A dominant ~22-kDa anti-His_5-reactive product was observed for rBfrB from both *E. coli* and *M. smegmatis*, and was close in size to the expected MWs of 22.0 and 21.5 kDa, respectively (Fig 4.2C). A dominant ~40 kDa anti-His_5-reactive product was observed for SP-rLppZ from *E. coli*, and was close in size to the expected MW of 37.6 kDa (Fig. 4.2D). More complex profiles were observed for both SP-rSodC from *E. coli* and rSodC from *M. smegmatis*, and these profiles represent glycosylated products, breakdown products, dimers, and multimers (Fig 4.2E) that are described in depth in chapter 5. Coomassie-stained protein

bands corresponding to the major anti-His_5-reactive product for each recombinant protein were digested with trypsin and subjected to LC-ESI-MS/MS. The resulting data were searched against all of the protein sequences extracted from the *Mtb* genome database (NC_000962), and in each case the expected protein identity was confirmed (data not shown).

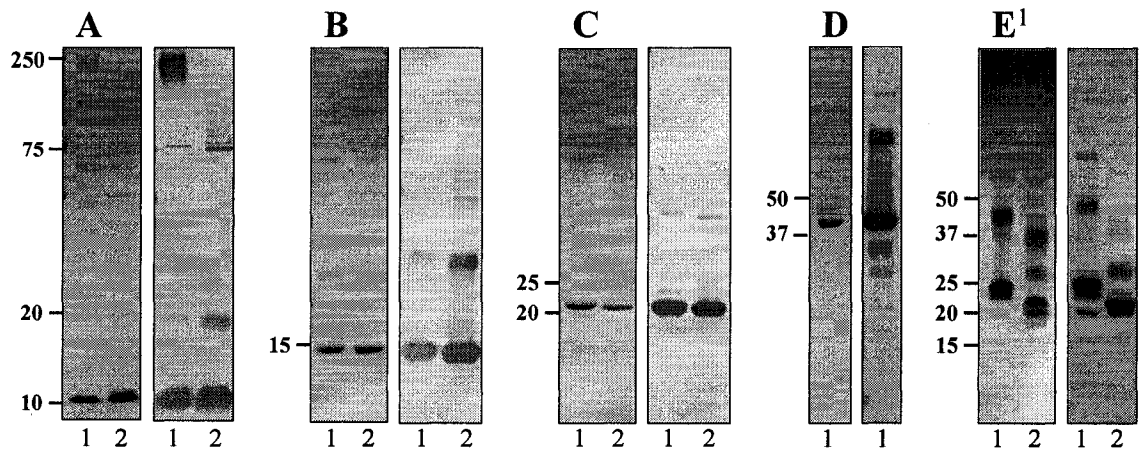


Fig. 4.2 Recombinant antigen production. The purities of recombinant proteins purified from *E. coli* (lane 1) or *M. smegmatis* (lane 2) were analyzed by SDS-PAGE and coomassie staining (panel 1), and Western blot using an anti-His monoclonal antibody as the probe (panel 2). (**A**) SecE2/Rv0379, (**B**) TrxC/Rv3914, (**C**) BfrB/Rv3841, (**D**) LppZ/Rv3006, (**E**) SodC/Rv0432. [1]SodC was stained with silver nitrate.

4.3.3 Subcellular localization

In order to assess the subcellular location of the native forms of the five antigens in *M. tuberculosis*, antibodies specific for each were developed. The purified recombinant proteins were used to immunize rabbits in a standard rabbit protocol. Western blot analyses showed the ability of each antiserum to recognize the respective recombinant antigen used for immunization (data not shown). *M. tuberculosis* H37Rv cytosol, membrane, cell wall, and CF fractions were isolated, separated by SDS-PAGE (Fig. 4.3A), electroblotted, and probed with polyclonal sera raised against SP-rLppZ

(Fig.4.3B), SP-rSodC (Fig. 4.3C), rBfrB (Fig. 4.3D), and rSecE2 (Fig. 4.3E). Each antiserum recognized a single dominant product of the expected MW, except for the anti-TrxC antiserum which exhibited significant cross-reactivity and is not included in figure 4.3. The LppZ and SodC products were found dominantly in the cell wall and membrane fractions, the expected subcellular location for secreted lipoproteins. In addition, LppZ and SodC products were detected in the CF. These products could represent proteolytically processed forms, a notion consistent with observations for other *Mtb* lipoproteins (13, 19). The BfrB products were found most dominantly in the membrane fraction, but were also observed in the cytosol and CF. The presence of BfrB in the membrane was unexpected, because BfrB is predicted to be a soluble, cytoplasmic protein; however, BfrB may be associated with protein complexes, perhaps forming interactions with membrane-bound iron transporters, although such transporters have yet to be confirmed in *Mtb* (23). BfrB lacks a putative signal peptide sequence and its presence in the CF is unexplained. Yet numerous other mycobacterial proteins lacking signal peptides have been found in the CF (3), raising the question as to whether these extracellular forms are due to 1) artifacts resulting from *in vitro* culture and experimental technique, or 2) an unidentified, active secretion process. The SecE2 products were found in all four fractions, and at least two reasons may explain this distribution. Although this protein lacks a signal peptide, it is annotated as a possible component of the Sec-dependent protein translocation apparatus (6), and this functional role may itself explain its presence in membrane and cell wall fractions, with the additional presence of SecE2 in the CF a direct result of cell wall turnover or leakage. Alternatively, the unusual physical properties of SecE2 (small size of 8.0 kDa, basic pI of 9.1, and

moderate hydrophobicity) may prevent its proper partitioning between the cytosol, membrane, and cell wall fractions with the subcellular fractionation methodology employed. In any case, the observed presence of fully secreted forms of LppZ, SodC, BfrB, and SecE2 in the CF is directly supportive of the protein identifications made in section 4.3.1.

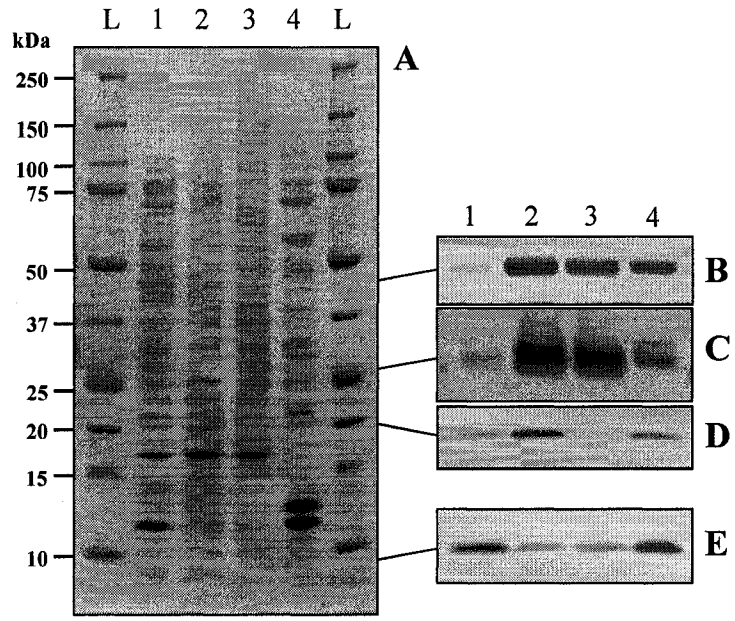


Fig. 4.3 Localization of antigens in *M. tuberculosis*. Subcellular fractions were separated by SDS-PAGE and stained with (**A**) Coomassie Brilliant Blue or (**B**) electroblotted and probed with polyclonal LppZ antiserum, (**C**) SodC antiserum, (**D**) TrxC antiserum, (**E**) or SecE2 antiserum. Lanes L, molecular weight marker, lanes 1, cytosol; lanes 2, membrane; lanes 3, cell wall; lanes 4, CF.

4.3.4 Antigen validation

The reactivities of serum samples from 37 cavitary-TB patients, 20 HIV+TB+ patients, 16 household contacts of untreated, infectious TB patients (HIV-TB- HH contacts), and 20 healthy controls (PPD+ and PPD-) with each of the purified recombinant proteins were determined by plate ELISA. The reactivities with

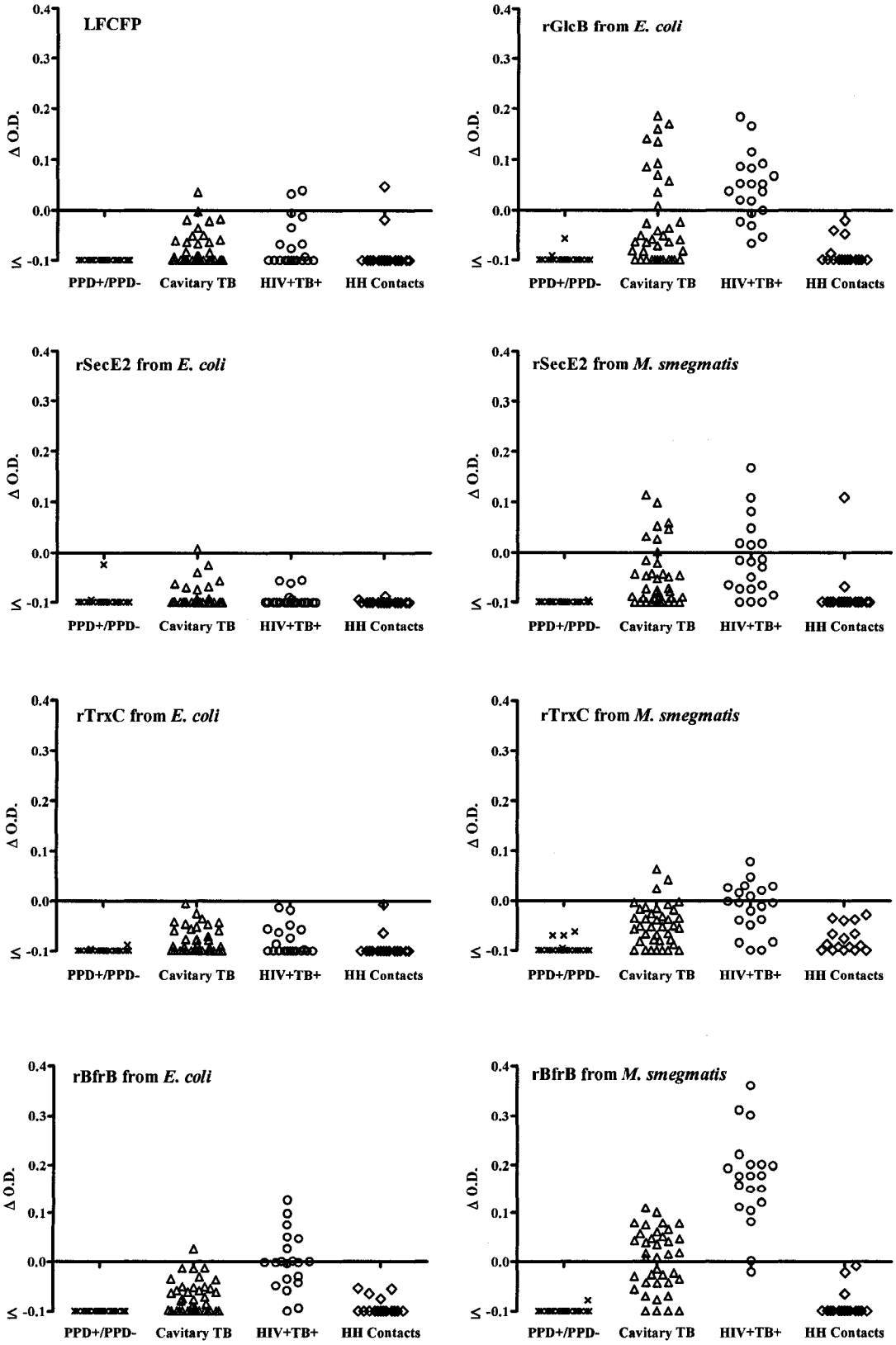
rGlcB/Mtb81 from *E. coli* and LFCFP, which served as positive controls, were also determined. There were no differences in the reactivities of sera obtained from PPD+ or PPD- healthy individuals for any of the antigens tested, suggesting that antibodies to these proteins are absent in healthy individuals, including those with latent, inactive *Mtb* infection or those BCG-vaccinated. To determine positive responses the mean OD obtained with the PPD+ and PPD- healthy individuals plus three SD was used as the cutoff, and representative single ELISA experiments are shown in Figure 4.4. Recombinant forms of TrxC and SecE2 initially displayed the poorest reactivity and were not tested further in the interest of sera conservation. For LFCFP and recombinant forms of BfrB, SodC, LppZ, and GlcB, each serum specimen was tested in three separate ELISA, and only sera which tested positive at least two times were considered positive. These results are summarized in Table 4.4. None of the 20 PPD+ and PPD- healthy control serum samples were reactive with any of the antigens, providing a specificity of 100% for each antigen.

Of the 37 smear-positive cavitary TB patients, serum samples from 22% (8 out of 37) and 3% (1 out of 37) of patients had antibodies to rSecE2 from *M. smegmatis* and *E. coli*. Serum samples from 8% (3 out of 37) and 0% (0 out of 37) of patients had antibodies to rTrxC from *M. smegmatis* and *E. coli*. Serum samples from 41% (15 out of 37) and 3% (1 out of 37) of patients had antibodies to rBfrB from *M. smegmatis* and *E. coli*, and serum samples from 32% (12 out of 37) and 5% (2 out of 37) of patients had antibodies to rSodC from *M. smegmatis* and SP-rSodC from *E. coli*, respectively. Thus, all four of the antigens provided higher sensitivities when produced in *M. smegmatis* versus *E. coli*. For this same group of patients, serum samples from 38% (14 out of 37)

of patients had antibodies to SP-rLppZ from *E. coli*, serum samples from 30% (11 out of 37) of patients had antibodies to rGlcB from *E. coli*, and serum samples from 3% (1 out of 37) of patients had antibodies to LFCFP.

In the cohort of HIV+TB+ patients, serum samples from 35% (7 out of 20) and 0% (0 out of 20) of patients had antibodies to rSecE2 from *M. smegmatis* and *E. coli*. Serum samples from 40% (8 out of 20) and 0% (0 out of 20) of patients had antibodies to rTrxC from *M. smegmatis* and *E. coli*. Serum samples from 90% (18 out of 20) and 30% (6 out of 20) of patients had antibodies to rBfrB from *M. smegmatis* and *E. coli*, and serum samples from 30% (6 out of 20) and 40% (8 out of 20) of patients had antibodies to rSodC from *M. smegmatis* and SP-rSodC from *E. coli*. Thus, three out of the four antigens (all but rSodC) provided higher sensitivities when produced in *M. smegmatis* versus *E. coli*. For this same group of patients, serum samples from 60% (12 out of 20) of patients had antibodies to SP-rLppZ from *E. coli*, serum samples from 75% (15 out of 20) of patients had antibodies to rGlcB from *E. coli*, and serum samples from 10% (2 out of 20) of patients had antibodies to LFCFP.

In the cohort of HIV-TB- HH contacts, serum samples from 6% (1 out of 16) of individuals had antibodies to both rSecE2 from *M. smegmatis* and LFCFP, and serum samples from 31% (5 out of 16) of individuals had antibodies to SP-rLppZ from *E. coli*. Interestingly, the same individual recognized all three of these antigens. The remaining antigens were not recognized by any of the contacts.



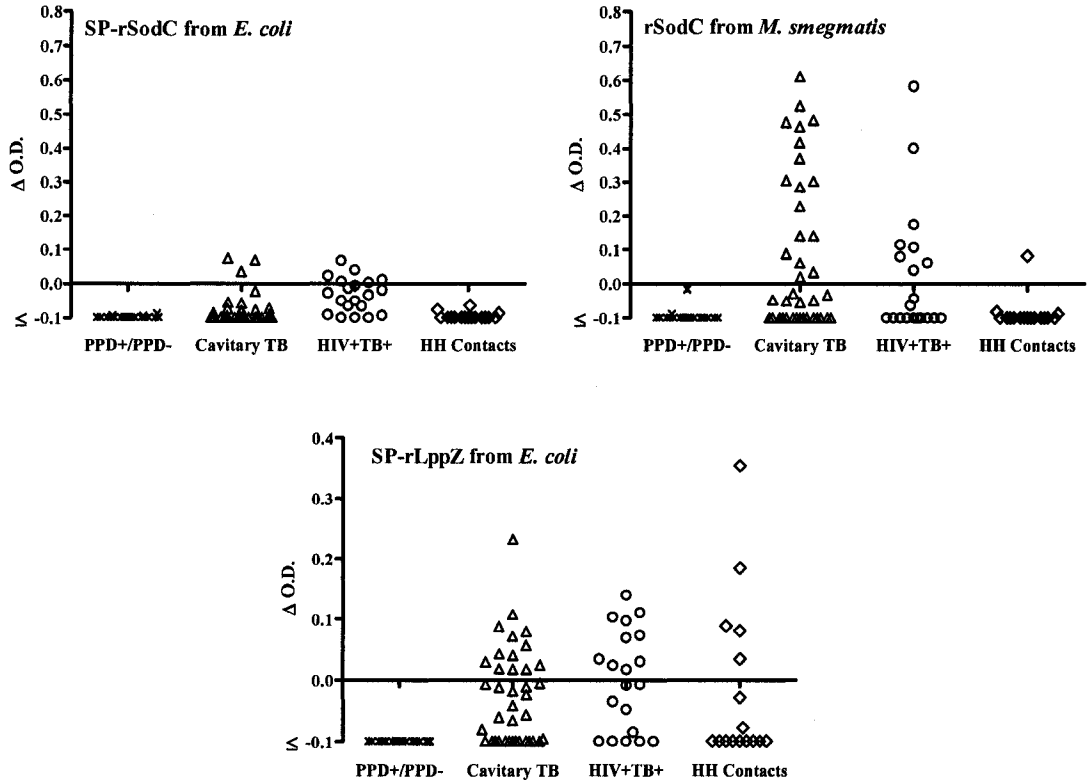


Fig. 4.4 Representative ELISA reactivities of purified recombinant proteins and LFCFP with human sera. Serum samples from healthy PPD+/PPD- healthy controls (**X**), Cavitary HIV-TB+ patients (Δ), HIV+TB+ patients (○), and HIV-TB-healthy-household contacts (◊) were tested by ELISA for reactivity against each antigen. The mean OD_{490} obtained with PPD+ and PPD- healthy individuals plus 3 SD for each antigen was used as the cutoff. The OD obtained with any serum specimen minus the cutoff (Δ O.D.) for each specimen is plotted, and responses above the horizontal line are considered positive for reactivity.

Table 4.4 Proportion of sera from TB patients and contacts containing antibodies to recombinant proteins and LFCFP[a]

Disease Status (*n*)	% of specimens containing antibodies to[b]:										
	E. coli rSecE2	*M. smeg* rSecE2	*E. coli* rTrxC	*M. smeg* rTrxC	*E. coli* rBfrB	*M. smeg* rBfrB	*E. coli* SP-rSodC	*M. smeg* rSodC	*E. coli* SP-rLppZ	*E. coli* rGlcB	LFCFP
Cavitary TB (37)	3 (1/37)	22 (8/37)	0 (0/37)	8 (3/37)	3 (1/37)	41 (15/37)	5 (2/37)	32 (12/37)	38 (14/37)	30 (11/37)	3 (1/37)
HIV+TB+ (20)	0 (0/20)	35 (7/20)	0 (0/20)	40 (8/20)	30 (6/20)	90 (18/20)	40 (8/20)	30 (6/20)	60 (12/20)	75 (15/20)	10 (2/20)
Healthy household contacts (16)	0 (0/16)	6 (1/16)	0 (0/16)	0 (0/16)	0 (0/16)	0 (0/16)	0 (0/16)	0 (0/16)	31 (5/16)	0 (0/16)	6 (1/16)
Specificity[c]	100	100	100	100	100	100	100	100	100	100	100

[a]The presence of antibodies to antigens in sera from TB patients and PPD+/PPD- healthy controls was determined by ELISA. The mean OD + 3SD obtained with the healthy controls was used as the cutoff to determine positive responses in patients

[b]Data in parentheses are the number of sera containing antibodies to the indicated antigen/the number of specimens tested. rBfrB, SP-rSodC, rSodC, SP-rLppZ, rGlcB, and LFCFP were tested in 3 separate ELISA, and positive responses were considered sera which tested positive 2 of 3 or 3 of 3 times. The rSecE2 and rTrxC responses are the result of a single ELISA

[c]Percentage of the 20 PPD+/PPD- healthy controls lacking antibodies to the indicated antigen.

The additive reactivity of sera with different antigen combinations was computed from the above data (Fig. 4.5). The combined antigen sensitivity of rSecE2, rTrxC, rBfrB, and SP-rSodC produced in *E. coli* (ALL-ECOLI) versus the combined antigen sensitivity of rSecE2, rTrxC, rBfrB, and rSodC produced in *M. smegmatis* (ALL-SMEG) was 11% and 62% for cavitary TB patients, and was 50% and 90% for HIV+TB+ patients. For both cohorts, each patient's serum sample recognized by ALL-ECOLI was also recognized by ALL-SMEG, meaning ALL-ECOLI was a complete subset of ALL-SMEG. Interestingly, for the HIV+TB+ patient cohort rBfrB from *M. smegmatis* alone provided a sensitivity of 90%, and this sensitivity could not be improved upon by combining the additive reactivity of any other antigen, including SP-rLppZ from *E. coli* and rGlcB from *E. coli*. For cavitary TB patients, combining the reactivity of ALL-SMEG with SP-rLppZ from *E. coli* raised the sensitivity to 73%, combining ALL-SMEG with rGlcB from *E. coli* raised the sensitivity to 65%, and combining ALL-SMEG with both SP-rLppZ from *E. coli* and rGlcB from *E. coli* raised the maximum sensitivity to 76%.

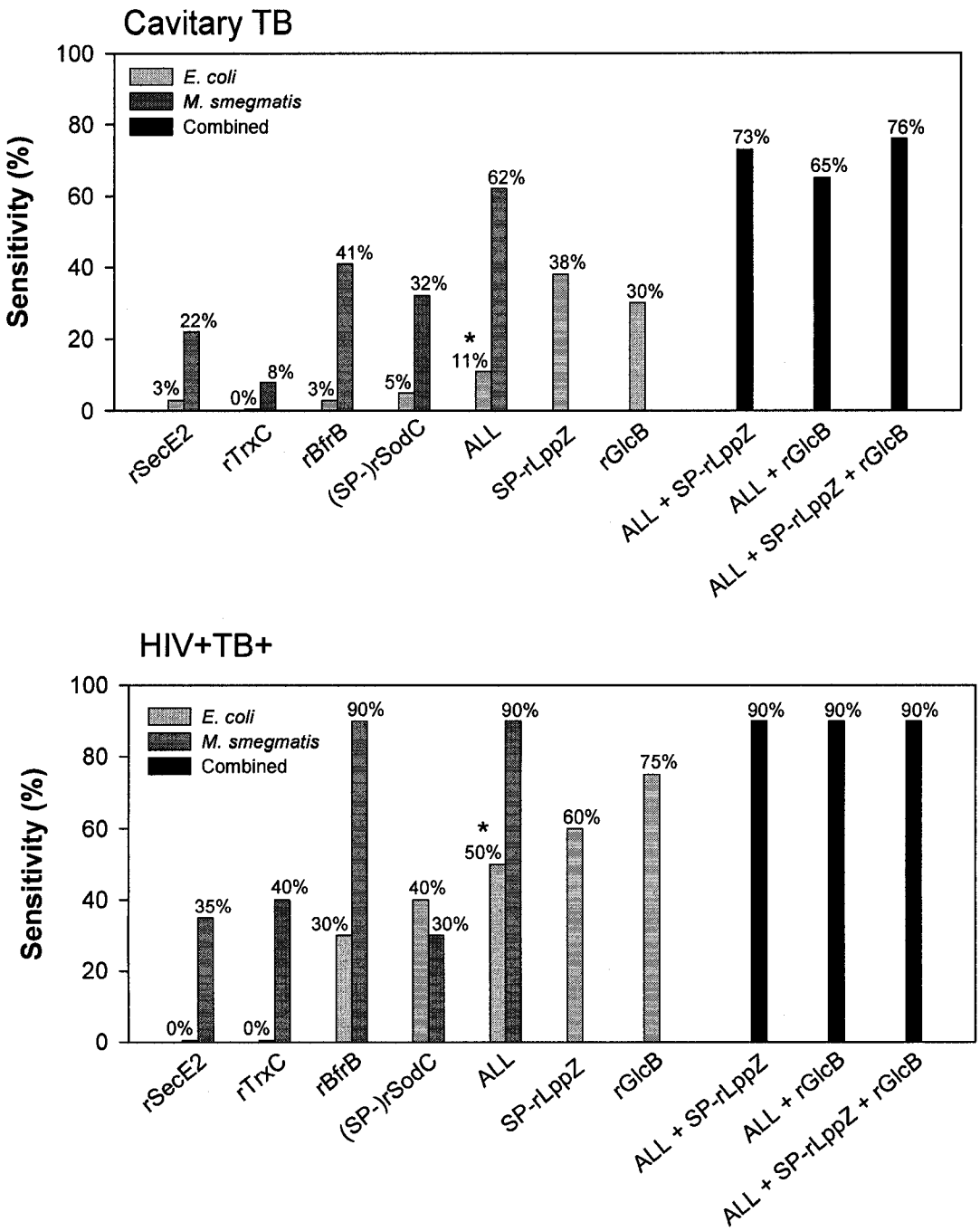


Fig. 4.5 Sensitivities of recombinant antigen combinations with cavitary TB and HIV+TB+ sera. The sensitivities of single recombinant proteins taken from Table 4.4 and the additive sensitivities of multiple antigen combinations are plotted. *ALL*, combined sensitivity of rSecE2 + rTrxC + rBfrB + (SP-)rSodC. The asterisk indicates that ALL from *E. coli* is a subset of ALL from *M. smegmatis*.

4.3.5 Protein microarray analyses identified antibody cross-specificity

In the antibody-based antigen identification approach described in section 3.3.4, two antibody preparations specific for LAM, the mAb CS-35 and a polyclonal LAM antisera, were initially used to probe array slides. Both antibodies reacted strongly against a series of spots corresponding to fractions from the initial flow-through of the RP column, and these fractions were confirmed to contain LAM as the dominant reactive species. Unexpectedly, the polyclonal LAM antisera was also reactive against an additional series of spots not recognized by CS-35 (Fig. 4.6A). These spots consistently corresponded to a single fraction collected from the RP column, and this fraction was also recognized significantly by TB patients' sera. When both the polyclonal LAM antisera and pooled cavitary-TB sera were used to probe these fractions on a Western blot, a single dominant ~10-kDa reactive product was observed rather than the expected 30- to 40-kDa broad band typically observed for LAM (Fig. 4.6A). Combined MS/MS analyses of peptides generated by trypsin, chymotrypsin, and endoproteinase Glu-C digestion of the corresponding 10-kDa coomassie-stained protein band in fraction B-I-4 resulted in 73% amino-acid coverage of the predicted protein sequence encoded by ORF *rv0379/secE2* (Table 3.3).

The differing microarray reactivity profiles of two antibody preparations thought to be specific for a single antigen were unexpected and required further investigation. To evaluate antibody specificity both CS-35 and the LAM antisera were used to probe purified LAM and rSecE2 from *E. coli* on a western blot. CS-35 recognized only the purified LAM, while the antisera recognized both LAM and rSecE2 from *E. coli* (Fig. 4.6B). The ability of the polyclonal LAM antisera to recognize rSecE2 purified from *E.*

coli provides strong evidence that the antibodies are binding SecE2 epitopes and not background mycobacterial molecules associating with the SecE2 protein. Interestingly, antisera raised against rSecE2 from *E. coli* recognized two ~6- and 7-kDa products in the purified LAM sample (Fig. 4.6C). These products likely correspond to proteolytic products of native SecE2 protein remaining following the pronase treatment used in the LAM purification procedure. Thus, the most plausible explanation for cross-reactivity of the polyclonal LAM antisera is that antibodies were raised against SecE2 polypeptides in addition to LAM in the initial immunization process.

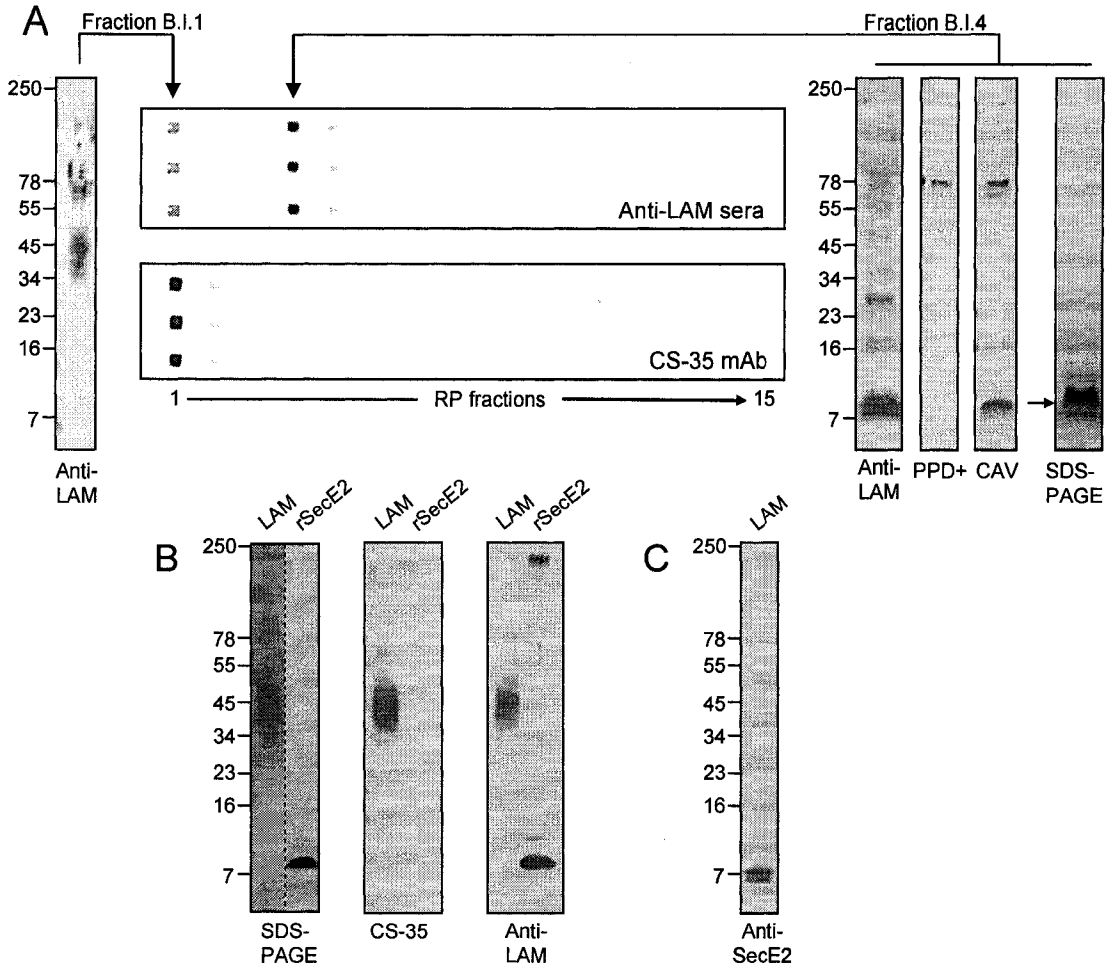


Fig. 4.6 Characterization of LAM antisera. Protein microarrays were probed with mAb CS-35 or polyclonal LAM antisera to identify fractions containing LAM as the dominant seroreactive antigen. Shown in (**A**) are portions of microarray slides containing ordered fractions eluting from a reversed phase column (printed in triplicate in the Y axis) and probed with CS-35 or LAM antisera. The microarray fractions recognized by the LAM antisera were further analyzed by western blot, and the LAM antisera recognized a ~34- to 46-kDa smear corresponding to LAM in fraction B.I.1 and a ~10-kDa product in fraction B.I.4 which was also recognized by TB patients' sera. MS/MS-based methods identified the corresponding 10-kDa coomassie-stained product (marked with →) as SecE2/Rv0379. (**B**) Purified LAM isolated from *Mtb* and recombinant SecE2 produced in *E. coli* were analyzed by SDS-PAGE, stained with silver nitrate or coomassie, respectively, and probed with CS-35 and LAM antisera on a western blot. (**C**) LAM was analyzed by western blot with antisera raised against recombinant SecE2 purified from *E. coli*.

4.4 Discussion

Over the past two decades molecular identification of the dominant B and T cell protein antigens of *Mtb* was achieved through a myriad of techniques and approaches (2, 18, 22). A major aim of this study was to develop protein separation strategies that would ultimately allow for enrichment of antigens present in low abundance. Combining this enrichment with newly emerging protein microarray technology, we were able to identify four potential novel serodiagnostic antigens previously undetected by other methods. Two of these antigens, SodC and LppZ, are believed to undergo post-translational modification. SodC was experimentally shown to be lipid-modified and associated with the bacterial membrane (9). This protein also contains three predicted glycosylation sites, and the 40-aa N-terminal fragment of SodC was found to bind ConA when fused to the 19-kDa-leader sequence and expressed in *M. smegmatis* (12). A method for predicting gram-positive lipoprotein motifs also identified LppZ as a probable lipoprotein (34), and similar to SodC, this protein sequence also contains three predicted glycosylation sites (12). One of the novel B cell antigens described in this work (BfrB) was previously shown to be a dominant T cell antigen for mice experimentally infected with *Mtb* (8), and this current work reveals that BfrB is also antigenic in human disease. Thus, like many other *Mtb* antigens, BfrB is strongly recognized by both the cellular and humoral arms of the immune system and could be targeted as either a vaccine candidate or diagnostic tool. The final B cell antigen newly recognized in this work (TrxC) was originally identified by Nagai *et al.* (21) as MPT46, a major product of the *Mtb* CF. Further work established MPT46 as a thioreductase (37). Although this protein was previously studied, the use of protein array technology has provided the first evidence of

its antigenic potential. Understanding that a comprehensive high-throughput experiment should require target validation as a final step, this study confirmed and characterized each antigen's serological response across multiple disease states.

In an effort to independently validate the seroreactivity of the novel antigens identified in this study, recombinant forms of each candidate antigen were produced and evaluated by conventional plate ELISA. All four novel antigens were recognized by at least 40% of one of the two TB patient subgroups tested, the same criteria initially used to identify microarray fractions with significant serological reactivity. The presence of TB patient antibodies against each of these four antigens not only provides new candidates for serodiagnostic tests, but also confirms the validity of both the protein microarray methodology developed and the data interpretation described in chapter three. An accurate assessment of the sensitivity of a new serodiagnostic antigen requires multiple independent tests with very large numbers of patients' sera across multiple cohorts. This pilot study falls short in both respects and therefore should not be considered an exhaustive evaluation; however, several aspects of the results are encouraging and provide great confidence in the designation of each recombinant protein as a true antigen with serodiagnostic potential. None of the 20 healthy individuals' sera recognized any of the proteins in this study, providing a specificity of 100% in each case. This attribute not only fulfills a basic criterion of a successful serodiagnostic test, but is also typical of most TB recombinant antigens where specificities are routinely found to exceed 95% (17). Furthermore, a seemingly appropriate number of HIV-TB- HH contacts' sera recognized most antigens (0 to 6%), considering an estimated 6 to 29% of the contacts were estimated to be at risk of developing TB within six months to two years

of infection (5, 10, 11, 36). Interestingly, a much greater percentage of contacts' sera recognized SP-rLppZ from *E. coli* (31%), and future studies are needed to confirm this observation and address the potential of LppZ for diagnosis of subclinical TB. Direct comparisons of the rGlcB responses in this study with previous data are, to some extent, reassuring. Anti-GlcB antibodies were present in 75% of HIV+TB+ patients' sera in this study, agreeing well with the 79% sensitivity rate previously described for rGlcB from *E. coli* with a similar patient cohort (28). However, anti-GlcB antibodies were present in only 30% of cavitary TB patients' sera in this study, in contrast to the previously described 70-77% sensitivity rates for the same antigen with similar cohorts (28, 29). In general, the responses of cavitary TB patients' sera among both test antigens and control antigens in this study were poorer than typically observed for this laboratory. Specifically, the differences in optical densities between TB patients and healthy controls were lower than routinely measured, suggesting the robustness of the assay may be suboptimal. This scenario would also explain the poor reactivity of LFCFP in comparison to similar patient cohorts in a previous study (29). Yet the reduced efficiency of the assay may in fact be encouraging, because the sensitivities assigned to each of the recombinant antigens in this study may in fact be underestimated. An especially promising outcome of this study was the level of antibody recognition for rBfrB from *M. smegmatis*, which was consistently recognized by more patients' sera than rGlcB, an antigen currently considered a leading candidate for use in serodiagnostic assays (17).

The SecE2/Mtb8/Rv0379 protein was identified in four microarray fractions recognized by both cavitary- and noncavitary-TB sera. Interestingly, a recent report suggested SecE2 to be a good candidate for multiple-antigen based serodiagnostic assays,

because SecE2 was recognized by TB patient sera with otherwise poor immunological reactivity (15). Therefore, to further evaluate the serodiagnostic potential of SecE2, recombinant forms were produced and evaluated by ELISA. Overall rSecE2 gave the poorest response of any of the recombinant proteins evaluated in this study, consistently detecting less than 40% of TB patient antibodies. Furthermore, each patient's serum that recognized rSecE2 also recognized at least one other antigen, suggesting SecE2 would not make a good complementary serodiagnostic antigen. However, analyses of rSecE2 allowed us to identify the underlying cause for the differing microarray reactivity profiles initially observed when arrays were probed with two antibody preparations specific for LAM. Ultimately the polyclonal LAM antisera was confirmed to possess cross-specificity for the SecE2 protein, providing an explanation for the extra set of spots recognized by LAM antisera but not mAb CS-35. Although an unintended outcome of this study, the characterization of antibody specificity via protein microarray technology was achieved, demonstrating yet another utility for this powerful methodology.

Native and recombinant forms of several *Mtb* proteins have been directly compared, and it is becoming increasingly clear that recombinant forms produced in *E. coli* often lack modifications or conformational epitopes required for immunological recognition (24, 26, 39). Two of the antigens identified in this study, SodC and LppZ, are predicted to be both lipidated and glycosylated (9, 12, 34). The discovery of SodC and LppZ as B cell antigens adds to the growing list of post-translationally modified antigens of *Mtb* (3) and illustrates the need for a recombinant protein production system that mimics native protein structures. Therefore, this study attempted to directly compare the seroreactivity of recombinant antigens produced in the fast-growing mycobacterial

species *M. smegmatis* with those produced in the standard *E. coli* system. Interestingly, each of the recombinant antigens produced in *M. smegmatis* were consistently recognized by more TB patient sera than their *E. coli* counterparts. However, three of the antigens tested, SecE2, TrxC, and BfrB, lack recognizable signal peptides and are not known or predicted to possess any PTMs, consistent with their observed SDS-PAGE profiles. Furthermore, the use of urea in the purification process likely destroyed most mycobacterial-specific conformational epitopes. Therefore, the overall observed differences in antibody recognition are likely not due to antigenic PTMs or epitopes resulting specifically from protein production in *M. smegmatis*. Instead, a more plausible explanation for the differences in seroreactivity is the presence of background *M. smegmatis* seroreactive molecules present in each of these purified recombinant protein preparations. The standard yet rather crude single-step immobilized metal ion affinity chromatography (IMAC) used for recombinant protein purification often results in appreciable amounts of both protein and non-protein contaminants (4). Indeed, some minor protein species not recognized by the anti-His_5 antibody can be detected when each of the purified recombinants was analyzed by SDS-PAGE and silver-staining (data not shown). TB patient antibodies could potentially recognize *M. smegmatis* proteins and non-proteinaceous molecules sharing homologous epitopes with *Mtb* seroreactive molecules. In addition to proteins, a number of non-proteinaceous *Mtb* molecules have been shown to be recognized by patient antibodies, including TDM, sulfolipid I, 2,3 diacyltrehalose, 2,3,6 triacyltrehalose, and LAM (16, 25). Probing each of the recombinant proteins with the mAb CS-35 on a Western blot demonstrated LAM only in

the rSodC purified from *M. smegmatis*, which may specifically explain the differential recognition of (SP-)rSodC and rSodC in cavitary TB patients (data not shown).

Recognition of *M. smegmatis* contaminants by TB patient antibodies would seem to improve the sensitivity of a serodiagnostic test based on the limited patient subgroups analyzed in this study. However, these contaminants could be problematic in cases where healthy individuals have been exposed to environmental mycobacteria, a very common occurrence in developing countries (1). Future efforts with recombinant antigens produced in *M. smegmatis* should focus on firmly establishing the presence of and identifying these contaminants, and then assessing their sensitivity and specificity for TB disease across multiple cohorts.

4.5 Literature Cited

1. **Andersen, P. and T. M. Doherty**. 2005. The success and failure of BCG - implications for a novel tuberculosis vaccine. Nat Rev Microbiol **3**(8): 656-62.

2. **Andersen, P. and T. M. Doherty**. 2005. TB subunit vaccines--putting the pieces together. Microbes Infect **7**(5-6): 911-21.

3. **Belisle, J. T., M. Braunstein, I. Rosenkrands, and P. Andersen**, The Proteome of *Mycobacterium tuberculosis*, in Tuberculosis and the Tubercle Bacillus, S.T. Cole, Editor. 2005, ASM Press: Washington, D.C. p. 235-60.

4. **Bolanos-Garcia, V. M. and O. R. Davies**. 2006. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. Biochim Biophys Acta **1760**(9): 1304-13.

5. **Claessens, N. J., F. F. Gausi, S. Meijnen, M. M. Weismuller, F. M. Salaniponi, and A. D. Harries**. 2002. High frequency of tuberculosis in households of index TB patients. Int J Tuberc Lung Dis **6**(3): 266-9.

6. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S.**

Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393**(6685): 537-44.

7. **Coligan, J. E. e. a.**, in Current Protocols in Protein Science. p. 10.5.1.

8. **Covert, B. A., J. S. Spencer, I. M. Orme, and J. T. Belisle**. 2001. The application of proteomics in defining the T cell antigens of *Mycobacterium tuberculosis*. Proteomics **1**(4): 574-86.

9. **D'Orazio, M., S. Folcarelli, F. Mariani, V. Colizzi, G. Rotilio, and A. Battistoni**. 2001. Lipid modification of the Cu,Zn superoxide dismutase from *Mycobacterium tuberculosis*. Biochem J **359**(Pt 1): 17-22.

10. **Doherty, T. M., A. Demissie, J. Olobo, D. Wolday, S. Britton, T. Eguale, P. Ravn, and P. Andersen**. 2002. Immune responses to the *Mycobacterium tuberculosis*-specific antigen ESAT-6 signal subclinical infection among contacts of tuberculosis patients. J Clin Microbiol **40**(2): 704-6.

11. **Grzybowski, S., G. D. Barnett, and K. Styblo**. 1975. Contacts of cases of active pulmonary tuberculosis. Bull Int Union Tuberc **50**(1): 90-106.

12. **Herrmann, J. L., R. Delahay, A. Gallagher, B. Robertson, and D. Young**. 2000. Analysis of post-translational modification of mycobacterial proteins using a cassette expression system. FEBS Lett **473**(3): 358-62.

13. **Herrmann, J. L., P. O'Gaora, A. Gallagher, J. E. Thole, and D. B. Young**. 1996. Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from *Mycobacterium tuberculosis*. Embo J **15**(14): 3547-54.

14. **Hirschfield, G. R., M. McNeil, and P. J. Brennan**. 1990. Peptidoglycan-associated polypeptides of *Mycobacterium tuberculosis*. J Bacteriol **172**(2): 1005-13.

15. **Houghton, R. L., M. J. Lodes, D. C. Dillon, L. D. Reynolds, C. H. Day, P. D. McNeill, R. C. Hendrickson, Y. A. Skeiky, D. P. Sampaio, R. Badaro, K. P. Lyashchenko, and S. G. Reed**. 2002. Use of multiepitope polyproteins in serodiagnosis of active tuberculosis. Clin Diagn Lab Immunol **9**(4): 883-91.

16. **Julian, E., L. Matas, A. Perez, J. Alcaide, M. A. Laneelle, and M. Luquin**. 2002. Serodiagnosis of tuberculosis: comparison of immunoglobulin A (IgA) response to sulfolipid I with IgG and IgM responses to 2,3-diacyltrehalose, 2,3,6-triacyltrehalose, and cord factor antigens. J Clin Microbiol **40**(10): 3782-8.

17. **Laal, S. and Y. A. Skeiky**, Immune-Based Methods, in Tuberculosis and the Tubercle Bacillus, S.T. Cole, Editor. 2005, ASM Press: Washington, D.C. p. 71-83.

18. **Laal, S. and S. Y.A.**, Immune-Based Methods, in Tuberculosis and the Tubercle Bacillus, S.T. Cole, Editor. 2005, ASM Press: Washington, D.C. p. 71-83.

19. **Michell, S. L., A. O. Whelan, P. R. Wheeler, M. Panico, R. L. Easton, A. T. Etienne, S. M. Haslam, A. Dell, H. R. Morris, A. J. Reason, J. L. Herrmann, D. B. Young, and R. G. Hewinson**. 2003. The MPB83 antigen from *Mycobacterium bovis* contains O-linked mannose and (1-->3)-mannobiose moieties. J Biol Chem **278**(18): 16423-32.

20. **Morrissey, J. H.** 1981. Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. Anal Biochem **117**(2): 307-10.

21. **Nagai, S., H. G. Wiker, M. Harboe, and M. Kinomoto**. 1991. Isolation and partial characterization of major protein antigens in the culture fluid of *Mycobacterium tuberculosis*. Infect Immun **59**(1): 372-82.

22. **Reed, S. and Y. Lobet**. 2005. Tuberculosis vaccine development; from mouse to man. Microbes Infect **7**(5-6): 922-31.

23. **Rodriguez, G. M. and I. Smith**. 2003. Mechanisms of iron regulation in mycobacteria: role in physiology and virulence. Mol Microbiol **47**(6): 1485-94.

24. **Romain, F., C. Horn, P. Pescher, A. Namane, M. Riviere, G. Puzo, O. Barzu, and G. Marchal**. 1999. Deglycosylation of the 45/47-kilodalton antigen complex of *Mycobacterium tuberculosis* decreases its capacity to elicit in vivo or in vitro cellular immune responses. Infect Immun **67**(11): 5567-72.

25. **Sada, E., P. J. Brennan, T. Herrera, and M. Torres**. 1990. Evaluation of lipoarabinomannan for the serological diagnosis of tuberculosis. J Clin Microbiol **28**(12): 2587-90.

26. **Samanich, K. M., M. A. Keen, V. D. Vissa, J. D. Harder, J. S. Spencer, J. T. Belisle, S. Zolla-Pazner, and S. Laal**. 2000. Serodiagnostic potential of culture filtrate antigens of *Mycobacterium tuberculosis*. Clin Diagn Lab Immunol **7**(4): 662-8.

27. **Sambrook J.E., F. E. F., Maniatis T.**, Molecular Cloning: A Laboratory Manual. 1989, New York: Cold Spring Harbor Laboratory Press.

28. **Singh, K. K., Y. Dong, J. T. Belisle, J. Harder, V. K. Arora, and S. Laal**. 2005. Antigens of *Mycobacterium tuberculosis* recognized by antibodies during incipient, subclinical tuberculosis. Clin Diagn Lab Immunol **12**(2): 354-8.

29. **Singh, K. K., Y. Dong, L. Hinds, M. A. Keen, J. T. Belisle, S. Zolla-Pazner, J. M. Achkar, A. J. Nadas, V. K. Arora, and S. Laal**. 2003. Combined use of serum and urinary antibody for diagnosis of tuberculosis. J Infect Dis **188**(3): 371-7.

30. **Smith, P. K., R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson, and D. C. Klenk.** 1985. Measurement of protein using bicinchoninic acid. Anal Biochem **150**(1): 76-85.

31. **Snapper, S. B., R. E. Melton, S. Mustafa, T. Kieser, and W. R. Jacobs, Jr.** 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. Mol Microbiol **4**(11): 1911-9.

32. **Sonnenberg, M. G. and J. T. Belisle.** 1997. Definition of *Mycobacterium tuberculosis* culture filtrate proteins by two-dimensional polyacrylamide gel electrophoresis, N-terminal amino acid sequencing, and electrospray mass spectrometry. Infect Immun **65**(11): 4515-24.

33. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol **185**: 60-89.

34. **Sutcliffe, I. C. and D. J. Harrington.** 2004. Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. FEMS Microbiol Rev **28**(5): 645-59.

35. **Takayama, K., H. K. Schnoes, E. L. Armstrong, and R. W. Boyle.** 1975. Site of inhibitory action of isoniazid in the synthesis of mycolic acids in *Mycobacterium tuberculosis*. J Lipid Res **16**(4): 308-17.

36. **Verver, S., R. M. Warren, Z. Munch, M. Richardson, G. D. van der Spuy, M. W. Borgdorff, M. A. Behr, N. Beyers, and P. D. van Helden.** 2004. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. Lancet **363**(9404): 212-4.

37. **Wieles, B., S. Nagai, H. G. Wiker, M. Harboe, and T. H. Ottenhoff.** 1995. Identification and functional characterization of thioredoxin of *Mycobacterium tuberculosis*. Infect Immun **63**(12): 4946-8.

38. **Wu, C. H., J. J. Tsai-Wu, Y. T. Huang, C. Y. Lin, G. G. Lioua, and F. J. Lee.** 1998. Identification and subcellular localization of a novel Cu,Zn superoxide dismutase of *Mycobacterium tuberculosis*. FEBS Lett **439**(1-2): 192-6.

39. **Zanetti, S., A. Bua, G. Delogu, C. Pusceddu, M. Mura, F. Saba, P. Pirina, C. Garzelli, C. Vertuccio, L. A. Sechi, and G. Fadda.** 2005. Patients with pulmonary tuberculosis develop a strong humoral response against methylated heparin-binding hemagglutinin. Clin Diagn Lab Immunol **12**(9): 1135-8.

Chapter V

N-terminal clustering of the O-Glycosylation Sites in the *Mycobacterium tuberculosis* Lipoprotein SodC

5.1 Introduction

Several of the immunodominant antigens of *Mtb* are reported to be glycosylated based on their ability to bind the lectin ConA (12, 14, 16, 17). Yet the presence of glycoproteins in *Mtb* did not gain wide acceptance until a MS-based analysis of the 45/47-kDa MPT32/Apa protein (Rv1860) demonstrated four separate Thr residues each O-linked with a mannose, mannobiose, or mannotriose (11). Although glycosylation of the 45/47-kDa protein influences acquired host immune responses (22, 38), and mediates interaction with host C-type lectins (37), a general role for protein glycosylation in *Mtb* remains elusive. This, however, has not precluded the identification of additional glycoproteins of *Mtb*. Two adjacent Thr residues from the secreted antigen MPB83 (Mb2898) of *M. bovis* were shown by MS-based methods to be modified with a total of three mannose residues (30). Using alternative approaches others have reported the glycosylation of recombinant *Mtb* proteins produced in *M. smegmatis*. Most recently, a member of the *Mtb* PPE protein family (Rv3873) produced in *M. smegmatis* appeared to be glycosylated at its C-terminus (9). Hermann *et al.* (20) detected glycosylation of two Thr clusters in the recombinant 19-kDa LpqH lipoprotein (Rv3763) and further showed that glycosylation of this region protected the recombinant 19-kDa lipoprotein from proteolytic cleavage. Additional work by this same investigator identified four putative glycolipoproteins using recombinant gene fusions in combination with ConA-based analyses (19). One of these putative lipoglycoproteins was identified as Rv0342 or

SodC. Specifically, a 40-aa N-terminal fragment of SodC was determined to be glycosylated by virtue of its ability to bind ConA.

In Chapters 3 and 4 we employed native antigen array profiling to identify and characterize serodiagnostic proteins of *Mtb*. Through this work four novel antigens were discovered including the *rv0432* gene product SodC. SodC is one of two *Mtb* SODs. SodC is a membrane-associated lipoprotein of the Cu,Zn-dependent SOD family (8), while the second SOD (SodA/Rv3846) is a Mn,Fe-dependent SOD believed to be actively secreted in a SecA2-dependent manner (5). The recognition of SodC and SodA (10) by human antibodies provides compelling evidence that both enzymes are produced in natural infection. Furthermore, SodC and SodA are shown to contribute to the survival of *Mtb* in activated macrophages (35) and in mice (13), respectively, suggesting each has a role in the defense against the oxidative burst produced *in vivo* and that one SOD cannot completely compensate for the absence of the other. Moreover, *sodC* transcription is greatly upregulated upon infection of human macrophages, suggesting SodC plays an important in early infection when the NADPH oxidase level is expected to be high (8, 51). Most recently, the *Mtb* SodC three-dimensional structure was solved by x-ray crystallography, and the active site of this fully functional enzyme was found to be devoid of zinc, an unusual feature not found in any other known copper-containing SODs (46). In light of these observations a unique opportunity presents itself to characterize the glycosylation of a known *Mtb* enzyme with a defined 3-D structure and function, as well as a putative role in virulence.

We set forth to characterize the glycosylation of the CF form of SodC from *Mtb*. Recombinant forms were produced in *E. coli* and *Mtb*, and recombinant SodC produced

in *Mtb* was found to bind ConA. The use of site directed mutagenesis combined with MS of peptides from recombinant variants of the *Mtb* SodC allowed for identification of up to six sites of glycosylation within a 13 amino acid region of the mature N-terminus of SodC. Further glycosylation at each site was with one to three α-mannose residues. These finding were in agreement with the previous work of (19), but for the first time also demonstrated O-glycosylation of serine (Ser) residues of an *Mtb* protein. Further, these studies help define the nature of O-glycosylated regions of *Mtb* proteins and provide additional evidence that protein glycosylation in *Mtb* functions to regulate proteolytic processing.

5.2 Materials and methods

5.2.1 Bacterial growth and subcellular fractionation

Recombinant clones of *M. smegmatis* mc^2155 were selected on LB agar containing kanamycin (25 μg ml^{-1}). For isolation of WCL of *M. smegmatis* cells were grown in 2 L of GAS medium (48) containing kanamycin (25 μg ml^{-1}) at 37°C for 3 days with gentle shaking, and bacterial pellets were suspended in PBS and passed through a French press four times at 2000 psi.

Mtb strain H37Rv cells used for the generation of electrocompetent stocks were grown in Middlebrook 7H9 broth (Difco) supplemented with oleic acid-dextrose catalase (OADC, Difco) and 0.05% Tween80. Recombinant clones of *Mtb* H37Rv were selected on Middlebrook 7H11 medium (Difco) supplemented with OADC, hygromycin (100 μg ml^{-1}), and kanamycin (25 μg ml^{-1}). Growth of *Mtb* for recombinant protein purification was achieved by propagation in 2 L of GAS medium containing kanamycin (25 μg ml^{-1}) at 37°C with gentle shaking for 14 days. Growth of *Mtb* for isolation of subcellular

fractions was achieved by propagation in 14 L of GAS medium at 37°C with gentle shaking. Cells were harvested at 14 days of growth, and individual subcellular fractions of cytosol, membrane, cell wall, and CF were isolated as previously described (21, 45). Final protein concentrations were determined using the BCA protein assay (43).

E. coli strains TOP10 (Invitrogen) and XL10-Gold Ultracompetent (Stratagene) containing various recombinant plasmids were grown on LB agar containing kanamycin (25μg ml^{-1}).

5.2.2 Bacterial transformation

For preparation of electrocompetent *Mtb* H37Rv, a 1 L culture was grown to an OD_{600} of 0.6 and the cells harvested by centrifugation. After washing three times in a 10% ice-cold glycerol solution the bacterial pellet was suspended in 5 ml of cold 10% glycerol. An aliquot (90 μl) of cells was mixed with 0.5 μg plasmid DNA, incubated in a 0.1 cm gap electroporation cuvette (Invitrogen) for 10 min at room temperature, and electroporated (1.25 kV, 25 μF, 1000Ω) in a Gene Pulser (Bio-Rad). Cells were transferred to 5 ml 7H9 medium, allowed to recover overnight at 37°C with gentle shaking, and plated onto solid medium containing antibiotics.

Electroporation of *M. smegmatis* mc^2155 was achieved by the method of Snapper *et al.* (44).

5.2.3 Construction of recombinant plasmids

Recombinant plasmid constructs were created according to standard protocols (40). PCR amplifications of *rv0432* or *rv0432* gene fragments were performed with *PfuTurbo* DNA polymerase (Stratagene) using *Mtb* H37Rv genomic DNA as the template. Table 5.1 lists PCR primers used in these studies. All PCR products were first

cloned into pCR4Blunt-TOPO according to the manufacturer's protocols (Invitrogen), and subsequently recovered by restriction enzyme digestions that targeted the restriction site linker sequences designed into the primers (Table 5.1). The expression constructs pMRLB60 for recombinant production of SP-rSodC in *E. coli*, and the mycobacterial expression construct pMRLB61 for recombinant production of mature and full modified rSodC, are described in section 4.2.2. To generate a mycobacterial-expression construct for production of SP-rSodC, the *rv0432* gene minus the region encoding the N-terminal signal peptide was amplified using the primer pair SodC(-)SPF and SodCR2. The 627 bp fragment isolated from the pCR4Blunt intermediate plasmid construct by digestion with *Nde*I/*Hind*III was ligated into the *Nde*I/*Hind*III sites of pVV16 to generate pMRLB62.

Mycobacterial expression constructs that allowed for production of rSodC with altered sites of glycosylation were based on pMRLB61 with point mutations in *rv0432*. Point mutations were generated using the QuikChange™ II XL site-directed mutagenesis kit (Stratagene) according to the manufacturer's recommendations. To generate the construct for rSodC-Thr_{41}-Ala_{41} production the primer pair SodCT41A41F and SodCT41A41R was used resulting in pMRLB61.1. To generate the construct for rSodC-Thr_{45}-Ala_{45} production the primer pair SodCT45A45F and SodCT45A45R was used resulting in pMRLB61.2. To generate the construct for rSodC-Thr_{46}-Ala_{46} production the primer pair SodCT46A46F and SodCT46A46R was used resulting in pMRLB61.3. To generate the construct for rSodC-$Thr_{45}Thr_{46}$-$Ala_{45}Ala_{46}$ production the primer pair SodCTTAAF and SodCTTAAR was used resulting in pMRLB61.4. To generate the construct for rSodC-Thr_{51}-Ala_{51} production the primer pair SodCT51A51F and SodCT51A51R was used resulting in pMRLB61.5. To generate the construct for rSodC-

Thr_{131}-Ala_{131} production the primer pair SodCT131A131F and SodCT131A131R was used resulting in pMRLB61.6. All plasmid constructs were confirmed by nucleotide sequencing through Macromolecular Resources Facility, Colorado State University.

Table 5.1 Plasmids and primers used in this study

	Description/Sequence	Reference or Source
Plasmid		
pCR4Blunt-TOPO	cloning vector, Amp^r, Kan^r	Invitrogen
pET23b	*E. coli* expression vector, Amp^r	Novagen
pVV16	*Mycobacteria* expression vector, Kan^r, Hyg^r	Schulbach *et al.* (2001)
pMRLB60	*rv0432* gene fragment lacking signal peptide sequence in pet23b	Chapter 4
pMRLB61	Full-length *rv0432* gene in pVV16	Chapter 4
pMRLB61.1	Derivative of pMRLB61 encoding Rv0432 Thr_{41} to Ala_{41} substitution	This study
pMRLB61.2	Derivative of pMRLB61 encoding Rv0432 Thr_{45} to Ala_{45} substitution	This study
pMRLB61.3	Derivative of pMRLB61 encoding Rv0432 Thr_{46} to Ala_{46} substitution	This study
pMRLB61.4	Derivative of pMRLB61 encoding Rv0432 $Thr_{45}Thr_{46}$ to $Ala_{45}Ala_{46}$ substitution	This study
pMRLB61.5	Derivative of pMRLB61 encoding Rv0432 Thr_{51} to Ala_{51} substitution	This study
pMRLB61.6	Derivative of pMRLB61 encoding Rv0432 Thr_{131} to Ala_{131} substitution	This study
pMRLB62	*rv0432* gene fragment lacking signal peptide sequence in pVV16	This study
Primer		
SodC(-SP)F	5'-CATATGTGCTCGTCGCCGCAG	Chapter 4
SodCR1	5'-CTCGAGGCCGGAACCAATGAC	Chapter 4
SodC(+SP)F	5'-CATATGCCAAAGCCCGCCGAT	Chapter 4
SodCR2	5'-AAGCTTGCCGGAACCAATGAC	Chapter 4
SodCT41A41F	5'-GCCGCAGCACGCGTCT**GCC**GTTCCGGGTACCACGCC	This study
SodCT41A41R	5'-GGCGTGGTACCCGGAAC**GGC**AGACGCGTGCTGCGGC	This study
SodCT45A45F	5'-CTACAGTTCCGGGT**GCC**ACGCCGTCGATTTGGACC	This study
SodCT45A45R	5'-GGTCCAAATCGACGGCGT**GGC**ACCCGGAACTGTAG	This study
SodCT46A46F	5'-CTACAGTTCCGGGTACC**GCG**CCGTCGATTTGGACC	This study
SodCT46A46R	5'-GGTCCAAATCGACGG**CGC**GGTACCCGGAACTGTAG	This study
SodCTTAAF	5'-CTACAGTTCCGGGT**GCCGCG**CCGTCGATTTGGACC	This study
SodCTTAAR	5'-GGTCCAAATCGA**CGGCGC**GGCACCCGGAACTGTAG	This study
SodCT51A51F	5'-GCCGTCGATTTGG**GCC**GGATCGCCCGCGCC	This study
SodCT51A51R	5'-GGCGCGGGCGATCC**GGC**CCAAATCGACGGC	This study
SodCT131A131F	5'-CAACTCGGTTGCCCCC**GCC**GGCGGTGCGCC	This study
SodCT131A131R	5'-GGCGCACCGCC**GGC**GGGGGCAACCGAGTTG	This study

Restriction endonuclease sites are underlined, site-directed mutations are in bold

5.2.4 Recombinant protein purification

Recombinant production of SP-rSodC in *E. coli* is described in section 4.2.3. For production of rSodC and rSodC site-directed mutants in *Mtb*, the plasmids pMRLB61, pMRLB61.1, pMRLB61.2, pMRLB61.3, pMRLB61.4, pMRLB61.5, and pMRLB61.6 were electroporated into *Mtb* H37Rv. CF from 14-day (late-log) cultures of recombinant *Mtb* were concentrated, dialyzed against 10 mM ammonium bicarbonate, and dried completely by lyophilization (45). The samples were suspended in 8 ml Ni-NTA denaturing binding buffer (0.1M sodium phosphate buffer, 8M urea, 10 mM Tris-Cl pH 8.0) and incubated 1 hour with 1 ml of Ni-NTA His-bind resin (Novagen) equilibrated in Ni-NTA denaturing binding buffer. The recombinant proteins were purified by washing with 18 CV of Ni-NTA denaturing binding buffer, 18 CV of Ni-NTA denaturing wash buffer (0.1M sodium phosphate buffer, 8M urea, 10 mM Tris-Cl pH 6.3), 10 CV of 10 mM Tris buffer (pH 8.0), 10 CV of 0.5% ASB-14 in 10 mM Tris buffer (pH 8.0), and again 10 CV of 10 mM Tris buffer (pH 8.0) to remove detergent. Proteins were eluted from the column by the addition of 10 CV of Ni-NTA denaturing elution buffer (0.1M sodium phosphate buffer, 8M urea, 10 mM Tris-Cl pH 4.0). Each fraction was concentrated ten-fold and exchanged with 10 mM ammonium bicarbonate by ultrafiltration.

5.2.5 SDS-PAGE and Western blot analyses

Rabbit polyclonal serum to SodC was generated by Strategic Biosolutions (Ramona, CA) in a standard rabbit protocol. Purified SP-rSodC from *E. coli* was used as the antigen.

Aliquots of protein were subjected to SDS-PAGE using 10-20% Tricine gels (Invitrogen). Gels were stained with Coomassie Brilliant Blue R250 (6) or silver nitrate (31), or electroblotted to nitrocellulose membranes (Bio-Rad) as previously described (45). The nitrocellulose membranes were blocked with 3% nonfat milk in PBS (pH 7.2) for 2 h and exposed overnight to anti-His_5 monoclonal antibody (Qiagen) or SodC antiserum diluted 1:1000 or 1:2000, respectively, in 1% nonfat milk in PBS (pH 7.2). The blots were washed with PBS (pH 7.2) containing 0.1% Tween 20, probed for 1.5 h with alkaline phosphatase-conjugated anti-mouse IgG whole molecule (Sigma) or anti-rabbit IgG Fc fragment (Calbiochem) both diluted 1:2000 in 1% nonfat milk in PBS (pH 7.2), and washed extensively. Antigen-antibody complexes were visualized by color development with 5-bromo-4-chloro-indoyl-phosphatase-nitroblue tetrazolium substrate (Kirkegaard & Perry Laboratories, Gaithersburg, MD). For ConA analysis, membranes were blocked with 1% BSA in PBS (pH 7.2) for 2 h and incubated with one unit of peroxidase-conjugated ConA (Sigma) in PBS (pH 7.2) overnight at 4°C. Development was achieved with 5-bromo-4-chloronaphthol (Sigma) and H_2O_2.

5.2.6 Analytical protein methods

Proteolytic digestions were performed with 10 μg aliquots of the various rSodC products purified from *Mtb*. Digestions with chymotrypsin (Roche Applied Science) were carried out in 20 μl of 0.1 M ammonium bicarbonate (pH 7.9) at 25°C for 16 h with an enzyme to substrate (E:S) ratio of 1:20 (wt:wt). Digestions with thermolysin (Calbiochem) were identical to those with chymotrypsin except the temperature was increased to 37 °C. Digestions with modified trypsin (Roche) were carried out in 20 μl 5% acetonitrile in 0.1 M ammonium bicarbonate (pH 7.9) at 37°C for 16 h with an E:S

ratio of 1:20 (wt:wt). All proteolytic digestions were terminated by the addition of 2 µl of 10% TFA. The digests were dried under vacuum, suspended in 20 µl of 5% acetonitrile in 0.1% acetic acid, and aliquots (5 µl) were applied to a capillary (0.75 x 50 mm) C18 reversed phase (RP) column (Agilent Technologies, Santa Clara, CA). The peptides were eluted with an increasing linear gradient (5% to 64%) of acetonitrile in 0.1% acetic acid over 70 min using an Agilent 1200 series capillary HPLC system with a flow rate of 5 µl per min, and introduced directly into a ThermoFinnigan LTQ electrospray mass spectrometer (San Jose, CA) operated using Xcalibur software ver 2.0 SR2. Ionization of peptides was achieved with an electrospray needle setting of 4 kV with a N_2 sheath gas flow of 15, and a capillary temperature of 200°C.

Multiple MS/MS methods were employed to analyze SodC peptides and their modifications. In all cases, data dependent scanning was used to generate fragment ions and MS/MS spectra. Standard MS/MS analyses were performed at 35% normalized collision energy of the top five most intense parent ions of the previous MS scan. Peptide identification was achieved using the SEQUEST (ThermoFinnigan, ver BioWorks 3.1) and X! Tandem (www.thegpm.org; version 2006.04.01.2) software with a SodC protein sequence (accession NP_214946) extracted from the *Mtb* genome database (NC_000962) and modified to begin with the experimentally determined N-terminal Thr_{41} and to include the additional C-terminal sequence of $_{241}KLHHHHHH_{248}$ (vector LysLeu + hexa-His tag). The SEQUEST and X! Tandem software were set to evaluate peptides obtained by chymotrypsin (FWYL), trypsin (KRLNH), or thermolysin (ILMV) digestion. Oxidation of Met (+16.0 Da) was set as the only possible modification. The Scaffold software, ver. Scaffold-01_05_04 (Proteome Software, Portland, OR) was used to

validate peptide identifications. Peptide identities were accepted only when a probability of 95.0% or greater was obtained as specified by the Peptide Prophet algorithm (41). MS/MS data were also searched for neutral losses of 162 *m/z* to identify glycopeptides. Neutral loss analyses of MS/MS spectra were performed with the Xcalibur software. ver 2.0 SR2.

To assess glycosylation of the Chy_1 peptide ($_{41}$TVPGTTPSIW$_{50}$) of rSodC the predicted $[M+H]^{+1}$ molecular ions of 1059.2, 1221.3, 1383.5, 1545.6, 1707.8, and 1869.9 *m/z* were placed in a parent mass list and when detected in a full MS scan, fragmentation for MS/MS was achieved using 21% normalized collision energy. The same approach was used for the Chy_1 peptides of rSodC-Thr$_{45}$-Ala$_{45}$ and rSodC-Thr$_{46}$-Ala$_{46}$ except the predicted $[M+H]^{+1}$ molecular ions of 1029.2, 1191.3, 1353.5, 1515.6, 1677.7, 1839.9 *m/z* were placed in the parent mass list. Glycosylation of the Chy_1 peptide of rSodC-Thr$_{41}$-Ala$_{41}$ ($_{42}$VPGTTPSIW$_{50}$) was assessed by placing the predicted $[M+H]^{+1}$ molecular ions of 958.1, 1120.2, 1282.4, 1444.5, 1606.7, and 1768.8 *m/z* in the parent mass list and performing MS/MS fragmentation with 21% normalized collision energy. The glycosylation analyses of the Chy_1 peptide of rSodC-Thr$_{45}$Thr$_{46}$-Ala$_{45}$Ala$_{46}$ ($_{45}$AAPSIW$_{50}$) utilized the predicted $[M+H]^{+1}$ molecular ions of 644.7, 806.9, 969.0, 1131.2, 1293.3, and 1455.4 *m/z* in the parent mass list.

MALDI-TOF MS was performed in the Macromolecular Resources Facility, Colorado State University, with an Ultraflex MALDI/TOF/TOF (Bruker Daltonics, Billerica, MA) with sinapinic acid as a matrix.

N-terminal sequencing of purified proteins was performed by Edman degradation (42) with a pulsed liquid-phase sequencer from Applied Biosystems by Macromolecular Resources Facility, Colorado State University.

5.2.7 Mannosidase treatment and Triton X-114 phase partitioning

Digestion of the protein with α-mannosidase was performed on aliquots (10 μg) of purified rSodC solubilized in 20 μl of 0.05 M sodium acetate buffer (pH 4.5) and incubated with 10, 2, 0.5, 0.1, and 0.02 μg of jack-bean-α-mannosidase from *Canavalia ensiformis* (Sigma) at 25°C for 18 h. Digestions were terminated by incubating with Laemmli sample buffer at 100°C for 10 min.

The detergent-phase partitioning properties of SodC were assessed by biphasic partitioning with Triton X-114 (Sigma-Aldrich). Specifically, subcellular fractions (CF, cell wall, and WCL) were incubated with 4% Triton X-114 in PBS (pH 7.4) for 16 h at 4°C with gentle agitation. The suspensions were incubated at 37°C for 1 h and the biphase was formed by centrifugation at 27,000 x *g* at 37°C for 1 h. The aqueous and detergent layers were collected, adjusted to 4% Triton X-114 in PBS (pH 7.4), and the biphasic partitioning repeated. Macromolecules in the final aqueous and detergent phases were precipitated by addition of 4 vol of ice-cold acetone and incubation at -20°C for 16 h. Precipitates were collected by centrifugation at 27,000 x *g* 4°C for 10 min, and washed once with ice-cold acetone. The precipitates were air-dried, suspended in PBS (pH 7.4), and protein concentrations were determined using the BCA protein assay (43).

5.2.8 Bioinformatics

The complete *Mtb* SodC protein sequence was analyzed with the NetOglyc 3.1 neural network for the prediction of O-linked glycosylation sites

(http://www.cbs.dtu.dk/services/NetOGlyc/) (24). Hydropathy plots were calculated for the mature *Mtb* SodC sequence (N-terminus of Thr_{41}) by the method of Kyte and Doolittle (27) via the ExPasy Proteomics Server (http://ca.expasy.org/tools/protscale.html) (18) using default parameters. To predict sequences which are intrinsically folded, the SodC sequence beginning with Cys_{33} was analyzed with the FoldIndex© tool (http://bip.weizmann.ac.il/fldbin/findex) (36) with a window size of 25. The non-mutated rSodC and rSodC-Thr_{41}-Ala_{41} proteins sequences were analyzed with the SignalP 3.0 server http://www.cbs.dtu.dk/services/SignalP/ with the Hidden Markov Model and Neural Network outputs for the prediction of signal peptide cleavage sites (32, 33). Protein sequences were aligned with Clustal W Version 1.83 software using default parameters (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_clustalw.html) (49).

5.3 Results

5.3.1 Expression of recombinant SodC in *E. coli* and *Mtb*

Chapter 4 described the identification of SodC as a potential serodiagnostic antigen that was present in the CF of *in vitro* grown *Mtb*, yet SodC was previously shown to be a lipoprotein presumably localized to the cellular envelope (8). Thus, the presence of this acylated product in the CF was unexpected. Further SodC is one of several *Mtb* lipoproteins that are suspected to be glycosylated (19). In an effort to further characterize this protein, recombinant SodC with a signal peptide (rSodC) and without a signal peptide (SP-rSodC) were produced in *Mtb* and *E. coli*, respectively. A recombinant form of SodC was produced in *Mtb* because quantities of the native product sufficient for biochemical analyses could not be purified from the CF. SP-rSodC was purified from the

cell lysate of *E. coli* and rSodC from the CF of *Mtb* by IMAC. Analysis of the recombinant products by SDS-PAGE revealed multiple products, regardless of source (Fig. 5.1A). The SP-rSodC from *E. coli* migrated at ~25 kDa, which is larger than the expected molecular weight (MW) of 21.8 kDa. This is a common observation for SODs in general (2), and has been specifically described for *Mtb* SodC produced in *E. coli* (8). Additionally, a putative dimer could be observed at ~47 kDa along with a series of multimers. The major rSodC product isolated from the *Mtb* CF migrated at ~28 kDa, with minor products at ~22 and 23 kDa, and a putative dimer at ~52 kDa. All of these products reacted with the anti-His_5 monoclonal antibody and SodC polyclonal antiserum. To ascertain the glycosylation status of SodC, the purified proteins were probed with the lectin ConA. The 28-kDa and 52-kDa rSodC products from *Mtb* displayed dominant binding to ConA, while the 22- and 23-kDa products clearly lacked ConA reactivity. As expected, the SP-rSodC from *E. coli* lacked ConA reactivity (Fig. 5.1A).

The presence of lower molecular mass rSodC products retaining reactivity against the anti-His_5 monoclonal antibody suggested breakdown is likely occurring at the N-terminus. Therefore, N-terminal amino acid sequencing of each product was performed to determine their full intact sequences. N-terminal sequencing of rSodC from *Mtb* revealed a sequence of $_{41}$TVPGXXPXI$_{49}$ for the dominant 28-kDa product, while sequencing of the two smaller products was more ambiguous with a mixture of products beginning with Gly_{58} and Ser_{60}. (Note, numbering of amino acids begins with $fMet_1$ of the full *sodC* gene product without processing.) The 25-kDa SP-rSodC product from *E. coli* produced an NH_2-terminus of $_{33}$CSSPQ$_{37}$, directly corresponding to the cloned gene fragment with processing of the fMet encoded by the ATG start codon (28, 52). These

N-terminal sequence data demonstrated the *Mtb* CF form of rSodC was not acylated. Additionally, the major product from *Mtb* was ~ 3 kDa greater in mass than that from *E. coli* yet was eight amino acids shorter. This indicated the presence of another modification on the *Mtb* product, presumably glycosylation. In support of this, jack-bean-α-mannosidase treatment of rSodC from *Mtb* resulted in a decrease in molecular mass (as observed by SDS-PAGE) and loss of ConA reactivity for the 28- and 52-kDa products (Fig. 5.1B). The magnitude of losses in mass and ConA-reactivity correlated with increasing concentrations of α-mannosidase. Thus, glycosylation appeared to significantly alter the mobility of SodC in SDS-PAGE. Treatment with α-mannosidase did not however have any effect on the migration lower mass (22- and 23-kDa) products (Fig. 5.1B). This with the lack of ConA reactivity indicated that the 22- and 23-kDa products lacked glycosylation and that glycosylation of the 28-kDa product occurred in the N-terminal region.

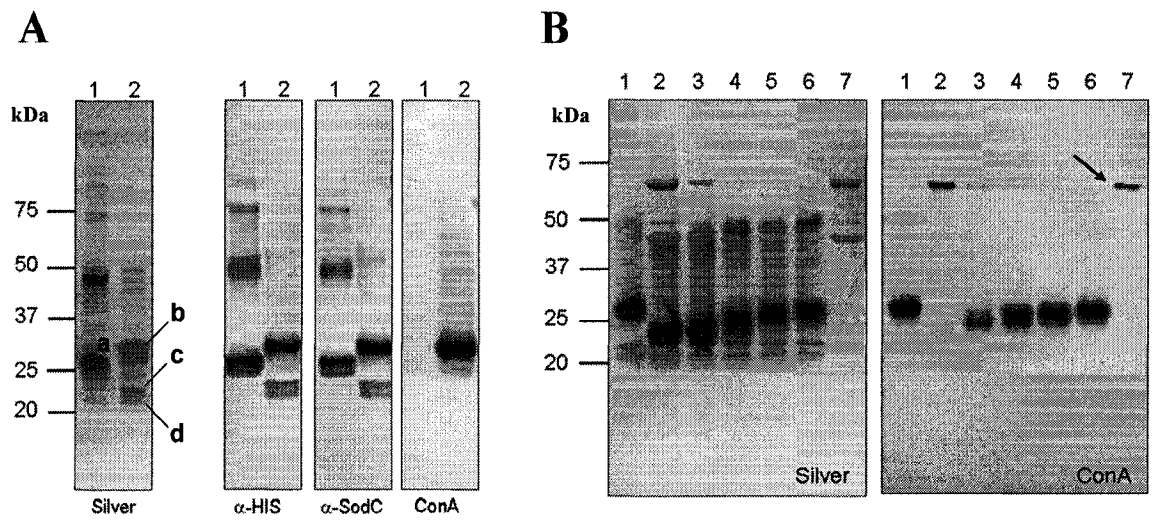


Fig. 5.1 ConA reactivity demonstrates glycosylation of *Mtb* rSodC. (**A**) Purified rSP-SodC (lane 1) and purified rSodC (lane 2) analyzed by SDS-PAGE and silver staining (panel 1) and Western blot analyses with anti-His_5 monoclonal antibody (panel 2), polyclonal SodC antiserum (panel 3), and ConA as probes (panel 4). The proteins displayed are: (a) the major 25-kDa product from *E. coli*; (b) the major 28-kDa product from *Mtb*; and (c and d) the minor 23- and 22-kDa products from *Mtb*. (**B**) Silver-stained SDS-PAGE (panel 1) and ConA Western blot (panel 2) analyses of purified rSodC from *Mtb* treated with various amounts α-mannosidase. Lane 1, untreated rSodC; lanes 2 to 6, rSodC (10 μg) treated with 10, 2, 0.2, 0.1 and 0.02 μg α-mannosidase; and lane 7, α-mannosidase alone. The 44- and 66-kDa subunits of α-mannosidase are observed in the silver stained gel (panel 1), and the 66-kDa subunit modified with a high-mannose-type-glycan (denoted by →) displays ConA reactivity (26).

5.3.2 MS-based identification of SodC glycopeptides

In order to identify SodC glycosylated regions(s), a variety of MS-based approaches were utilized. Purified rSodC from *Mtb* was digested with trypsin, chymotrypsin, and thermolysin and analyzed by LC-ESI-MS/MS, and the resulting MS/MS data searched using SEQUEST (29) and X! Tandem (15) software against the experimentally determined recombinant SodC protein sequence. As summarized in Figure 5.2, interrogation of the MS/MS data and peptide identification performed without inclusion of potential O-glycosylation resulted in the recombinant protein being mapped

in its entirety, except for the N-terminal Thr_{41}-Leu_{59} sequence and the peptide $_{123}CEPNS_{127}$ (Table 5.2). The inability to identify unmodified N-terminal peptide(s) was consistent with ConA reactivity and N-terminal sequencing results that indicate glycosylation of the N-terminal region.

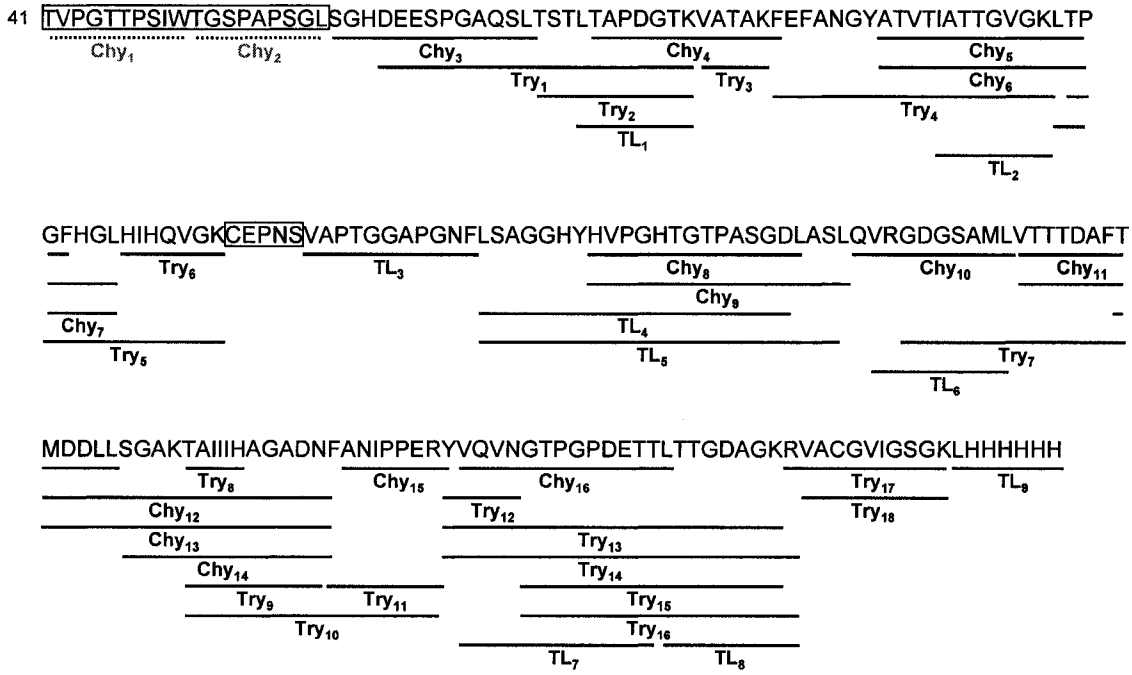


Fig. 5.2 Mapping SodC peptides. The deduced amino acid sequence (beginning with Thr_{41}) of the 28-kDa rSodC protein product purified from *Mtb* and alignment with the peptides generated by chymotrypsin, trypsin, and thermolysin digestion is shown. The solid lines indicate the location of individual unmodified peptides identified by LC-ESI-MS/MS, and the boxes indicate unidentified sequences. The dashed lines indicate two predicted chymotrypsin derived peptides used for the remainder of this work. *Chy*, chymotrypsin; *Try,* trypsin; *TL,* thermolysin

Table 5.2 MS-based identification of rSodC nonglycosylated peptides

Trypsin

Peptide	Sequence	Peptide Start Index	Peptide Stop Index	Previous Amino Acid	Next Amino Acid	Number of spectra	Spectrum Charge	Modifications	Peptide identification probability	Best SEQUEST XCorr score	Best X! Tandem -log(e) score	Precursor Mass Range (m/z)	Actual peptide mass range (AMU)	Actual minus calculated calculated peptide mass range (AMU)
Try1	DEESPGAQSLTSTLTAPDGTK	23	43	H	V	3	2		95%	5.55	0	1053.42 - 1053.86	2104.82 - 2105.71	0.84 - 1.7
Try2	TSTLTAPDGTK	33	43	L	V	4	1		95%	2.45	0	1091.68 - 1092.74	1090.68 - 1091.73	0.12 - 1.2
Try2	TSTLTAPDGTK	33	43	L	V	3	2		95%	2.67	0	546.74 - 547.23	1091.46 - 1092.44	0.91 - 1.9
Try3	VATAK	44	48	K	F	25	1		95%	2.3	0.699	488.87 - 490.66	487.86 - 489.65	-0.43 - 1.44
Try4	FEFANGYATVTIATTGVGK	49	67	K	L	13	1		95%	3.13	6.85	1947.99 - 1950.04	1946.98 - 1949.03	1 - 3.1
Try4	FEFANGYATVTIATTGVGK	49	67	K	L	131	2		95%	6.13	12.8	974.11 - 976.25	1946.2 - 1950.48	0.22 - 4.5
Try4	FEFANGYATVTIATTGVGK	49	67	K	L	66	3		95%	5.02	13.2	650.19 - 651.11	1947.55 - 1950.31	1.6 - 4.3
Try5	LTPGFHGLHIHQVGK	68	82	K	C	3	3		95%	0	6.59	547.6 - 548.53	1639.77 - 1642.56	-0.12 - 2.7
Try6	HIHQVGK	76	82	L	C	2	2		95%	2.41	0	409.91 - 410.44	817.81 - 818.87	0.36 - 1.4
Try7	GDGSAMLVTTTDAFTMDDLLSGAK	126	149	R	T	8	2	2 Met Oxidations	95%	0	9.64	1225.48 - 1226.36	2448.95 - 2450.7	0.85 - 2.6
Try7	GDGSAMLVTTTDAFTMDDLLSGAK	126	149	R	T	89	2	Met Oxidation	95%	0	11.9	1216.84 - 1219.06	2431.66 - 2436.1	-0.45 - 4
Try7	GDGSAMLVTTTDAFTMDDLLSGAK	126	149	R	T	110	2		95%	0	13.6	1208.86 - 1210.93	2415.7 - 2419.84	-0.41 - 3.7
Try7	GDGSAMLVTTTDAFTMDDLLSGAK	126	149	R	T	6	3	2 Met Oxidations	95%	0	12.8	817.31 - 818.04	2448.91 - 2451.09	0.81 - 3
Try7	GDGSAMLVTTTDAFTMDDLLSGAK	126	149	R	T	62	3	Met Oxidation	95%	0	14.2	811.85 - 813.1	2432.53 - 2436.27	0.42 - 4.2
Try7	GDGSAMLVTTTDAFTMDDLLSGAK	126	149	R	T	77	3		95%	0	14.9	806.41 - 807.75	2416.21 - 2420.21	0.098 - 4.1
Try8	TAIIIH	150	155	K	A	5	1		95%	2.18	0	667.56 - 668.64	666.55 - 667.63	0.15 - 1.2
Try8	TAIIIH	150	155	K	A	3	2		95%	2.27	0	334.58 - 334.65	667.14 - 667.29	0.74 - 0.89
Try9	TAIIIHAGADN	150	160	K	F	1	2		95%	3.54	0	548.67	1095.32	0.75
Try10	TAIIIHAGADNFANIPPER	150	168	K	Y	43	2		95%	0	15.5	1010.59 - 1012.03	2019.16 - 2022.04	0.11 - 3
Try10	TAIIIHAGADNFANIPPER	150	168	K	Y	115	3		95%	0	12.4	673.78 - 675.41	2018.32 - 2023.19	-0.73 - 4.1
Try11	FANIPPER	161	168	N	Y	1	2		95%	2.77	0	472.72	943.43	0.94
Try12	YVQVN	169	173	R	G	4	1		95%	1.84	0	622.51 - 623.52	621.5 - 622.51	0.19 - 1.2
Try13	YVQVNGTPGPDETTLTTGDAGK	169	190	R	R	5	2		95%	6.18	8.85	1111.56 - 1113.07	2221.12224.12	1.1 - 4.1
Try13	YVQVNGTPGPDETTLTTGDAGK	169	190	R	R	6	3		95%	13.5	14.75	741.83 - 742.42	2222.46 - 2224.23	2.4 - 4.2
Try14	YVQVNGTPGPDETTLTTGDAGKR	169	191	R	V	5	2		95%	0	9.96	1189.5 - 1190.51	2376.98 - 2379.01	0.83 - 2.9
Try14	YVQVNGTPGPDETTLTTGDAGKR	169	191	R	V	3	3		95%	0	7.5 - 7.77	793.53 - 793.87	2377.56 - 2378.58	1.4 - 2.4
Try15	GTPGPDETTLTTGDAGK	174	190	N	R	1	2		95%	2.96	0	810.1	1618.19	1.4
Try16	GTPGPDETTLTTGDAGKR	174	191	N	V	4	2		95%	3.36	0	887.8 - 888.95	1773.59 - 1775.89	0.74 - 3
Try17	RVACGVIGSGK	191	201	K	L	2	2		95%	2.39	1.15	524.05 - 524.08	1046.08 - 1046.15	0.51 - 0.58
Try18	VACGVIGSGK	192	201	R	L	3	2		95%	2.43	2.43	445.79 - 446	889.56 - 889.99	0.093 - 0.52

Chymotrypsin

Peptide	Sequence	Peptide Start Index	Peptide Stop Index	Previous Amino Acid	Next Amino Acid	Number of spectra	Spectrum Charge	Modifications	Peptide identification probability	Best SEQUEST XCorr score	Best X! Tandem -log(e) score	Precursor Mass Range (m/z)	Actual peptide mass range (AMU)	Actual minus calculated calculated peptide mass range (AMU)
Chy3	SGHDEESPGAQSL	20	32	L	T	3	1		95%	3.59	5.57	1314.13 - 1314.88	1313.12 - 1313.87	0.57 - 1.3
Chy3	SGHDEESPGAQSL	20	32	L	T	11	2		95%	3.23	4.75	657.56 - 658.68	1313.11 - 1315.35	0.56 - 2.8
Chy4	TAPDGTKVATAKF	37	49	L	E	4	1		95%	2.19	3.12	1306.86 - 1307.83	1305.85 - 1306.82	0.16 - 1.1
Chy4	TAPDGTKVATAKF	37	49	L	E	13	2		95%	3.28	4.77	653.25 - 655.28	1304.48 - 1308.54	-1.2 - 2.9
Chy4	TAPDGTKVATAKF	37	49	L	E	2	3		95%	1.11	5.36	436.65 - 436.85	1306.93 - 1307.52	1.2 - 1.8
Chy5	ATVTIATTGVGKLTPGF	56	72	Y	H	2	2		95%	2.69	5.28	817.87 - 818	1633.73 - 1633.99	0.82 - 1.1
Chy6	ATVTIATTGVGKLTPGFHGL	56	75	Y	H	1	2		95%	3.98	7.41	971.53	1941.05	0.98
Chy6	ATVTIATTGVGKLTPGFHGL	56	75	Y	H	3	3		95%	3.91	10.7	648.37 - 648.53	1942.09 - 1942.57	2 - 2.5
Chy7	TPGFHGL	69	75	L	H	3	2		95%	2.17	2.75	365.09 - 365.28	728.16 - 728.55	0.8 - 1.2
Chy8	HVPGHTGTPASGDL	106	119	Y	A	4	1		95%	2.39	4.68	1345.77 - 1346.8	1344.76 - 1345.79	0.12 - 1.2
Chy8	HVPGHTGTPASGDL	106	119	Y	A	12	2		95%	2.76	5.34	673.73 - 674.83	1345.45 - 1347.64	0.81 - 3
Chy8	HVPGHTGTPASGDL	106	119	Y	A	1	3		95%	1.37	4.3	449.86	1346.56	1.9
Chy9	HVPGHTGTPASGDLASL	106	122	Y	Q	2	1		95%	2.46	4.68	1617.34 - 1617.92	1617.92 - 1616.91	0.54 - 1.1
Chy9	HVPGHTGTPASGDLASL	106	122	Y	Q	7	2		95%	4.01	5.24	809.34 - 810.45	1616.67 - 1618.89	0.88 - 3.1
Chy9	HVPGHTGTPASGDLASL	106	122	Y	Q	6	3		95%	2.29	8.57	539.46 - 540.34	1615.36 - 1618	-0.43 - 2.2
Chy10	QVRGDGSAML	123	132	L	V	2	1		95%	2.23	2.46	1034.65 - 1034.72	1033.64 - 1033.71	1.1 - 1.2
Chy10	QVRGDGSAML	123	132	L	V	9	2	Met Oxidation	95%	2.71	4.46	525.63 - 526.69	1049.25 - 1051.37	0.76 - 2.9
Chy10	QVRGDGSAML	123	132	L	V	7	2		95%	2.02	4.25	517.58 - 518.56	1033.14 - 1035.11	0.64 - 2.6
Chy11	VTTTDAFTMDDLL	133	145	L	S	4	2	Met Oxidation	95%	3.47	4.02	730.33 - 731.28	1458.65 - 1460.54	0.99 - 2.9
Chy11	VTTTDAFTMDDLL	133	145	L	S	3	2		95%	3.69	5.43	722.27 - 722.4	1442.53 - 1442.79	0.87 - 1.1
Chy12	VTTTDAFTMDDLLSGAKTAIIIHAGADNF	133	161	L	A	2	3		95%	0	11.4	1004.53 - 1005.02	3010.56 - 3012.02	2.1 - 3.5
Chy13	TMDDLLSGAKTAIIIHAGADNF	140	161	F	A	2	2	Met Oxidation	95%	6.17	7.59	1146.01 - 1146.94	2290.01 - 2291.86	0.88 - 2.7
Chy13	TMDDLLSGAKTAIIIHAGADNF	140	161	F	A	11	2		95%	5.9	6.66	1137.91 - 1138.95	2273.81 - 2275.88	0.68 - 2.8
Chy13	TMDDLLSGAKTAIIIHAGADNF	140	161	F	A	65	3	Met Oxidation	95%	5.86	12.2	763.96 - 765.43	2288.84 - 2293.25	-0.29 - 4.1
Chy13	TMDDLLSGAKTAIIIHAGADNF	140	161	F	A	99	3		95%	5.61	14.3	758.64 - 760.22	2272.89 - 2277.64	-0.24 - 4.5
Chy14	SGAKTAIIIHAGADNF	146	161	L	A	124	2		95%	5.35	8.34	793.01 - 795.43	1584.01 - 1588.84	-0.81 - 4
Chy14	SGAKTAIIIHAGADNF	146	161	L	A	95	3		95%	3.97	6.75	529.24 - 530.39	1584.7 - 1588.14	-0.12 - 3.3
Chy15	ANIPPERY	162	169	F	V	21	2		95%	2.17	3.15	480.49 - 480.84	958.97 - 959.67	0.49 - 1.2
Chy16	VQVNGTPGPDETTL	170	183	Y	T	6	1		95%	2.47	4.12	1428.37 - 1429.9	1427.36 - 1428.89	0.67 - 2.2
Chy16	VQVNGTPGPDETTL	170	183	Y	T	5	2		95%	3.14	6.66	714.81 - 715.8	1427.6 - 1429.59	0.91 - 2.9

Thermolysin

Peptide	Sequence	Peptide Start Index	Peptide Stop Index	Previous Amino Acid	Next Amino Acid	Number of spectra	Spectrum Charge	Modifications	Peptide identification probability	Best SEQUEST XCorr score	Best X! Tandem -log(e) score	Precursor Mass Range (*m/z*)	Actual peptide mass range (AMU)	Actual minus calculated calculated peptide mass range (AMU)
TL1	LTAPDGTK	36	43	T	V	4	1		95%	1.42	1.89	802.46 - 802.8	801.45 - 801.79	0.029 - 0.37
TL2	IATTGVGK	60	67	T	L	1	1		95%	1.56	1.13	746.52	745.51	0.079
TL3	VAPTGGAPGNF	88	98	S	L	8	1		95%	2.56	4.39	987.5 - 989.53	986.49 - 988.52	0.01 - 2
TL3	VAPTGGAPGNF	88	98	S	L	4	2		95%	2.74	6.72	494.46 - 494.98	986.91 - 987.95	0.43 - 1.5
TL4	LSAGGHYHVPGHTGTPASGDLAS	99	121	F	L	4	2		95%	3.47	8.43	1095.3 - 1095.85	2188.59 - 2189.69	0.56 - 1.7
TL4	LSAGGHYHVPGHTGTPASGDLAS	99	121	F	L	5	3		95%	1.88	7.92	730.11 - 730.9	2187.3 - 2189.68	-0.73 - 1.7
TL5	LSAGGHYHVPGHTGTPASGD	99	118	F	L	2	3		95%	3.61	7.96	640.5 - 640.81	1918.48 - 1919.41	1.6 - 2.5
TL6	VRGDGSAML	124	132	Q	V	2	1		95%	1.93	2.08	905.63 - 905.7	904.62 - 904.69	0.18 - 0.25
TL7	VQVNGTPGPDETT	170	182	Y	L	1	1		95%	1.84	3.96	1315.56	1314.55	0.94
TL7	VQVNGTPGPDETT	170	182	Y	L	1	2		95%	2	6	658.56	1315.1	1.5
TL8	LTTGDAGKR	183	191	T	V	1	1		95%	1.74	2.85	918.56	917.55	0.059
TL8	LTTGDAGKR	183	191	T	V	6	2		95%	2.77	3.96	459.92 - 460.2	917.83 - 918.39	0.34 - 0.9
TL9	LHHHHHH	202	208	K	-	4	2		95%	2.69	2.28	477.99 - 478.19	953.96 - 954.36	0.91
TL9	LHHHHHH	202	208	K	-	1	3		95%	2.28	2.38	319.15	954.42	0.97

To identify glycopeptides the MS/MS data from the chymotrypsin digest were searched for neutral losses of 162 *m/z* (mass to charge ratio), a diagnostic property of the dissociation of hexosyl residues from a singly charged glycopeptide (11). Three major $[M+H]^{+1}$ molecular ions (1110.5, 1272.5, and 1434.6 *m/z*) were found that differed by 162 *m/z* and corresponded to the peptide designated Chy_2 ($_{51}$TGSPAPSGL$_{59}$) with two, three, and four hexosyl residues, respectively (Fig. 5.2 and 5.3A). Fragmentation and MS/MS of the largest molecular ion (1434.6 *m/z*) generated the $[M+H]^{+1}$ daughter ions of 1272.5, 1110.5, 948.5, and 786.5 *m/z* arising from neutral losses of 162 *m/z* (Fig. 5.3B). The 786.5 *m/z* fragmentation ion corresponds to the non-glycosylated peptide that has a predicted $[M+H]^{+1}$ mass of 786.9 *m/z*. A similar search of the MS/MS data for molecular ions and neutral losses corresponding to a Chy_1 glycopeptide were unsuccessful.

The inability to identify a Chy_1 glycopeptide was not unexpected since ionization suppression is a known phenomenon for ESI-MS-based analyses of glycopeptides in the presence of unmodified peptides (7). Therefore, a more targeted approach was designed, whereby molecular ions with *m/z* values corresponding to a predicted Chy_2 peptide modified with zero to five hexoses were selected from the full MS scan and further subjected to MS/MS analyses. Two major $[M+H]^{+1}$ molecular ions (1707.8 and 1869.8 and *m/z*) differing by 162 *m/z* and corresponding to the Chy_1 ($_{41}$TVPGTTPSIW$_{50}$) peptide with four and five hexosyl residues, respectively, were found in the MS data (Fig. 5.3C). MS/MS analysis of the larger molecular ion (1869.8 m/z) generated the $[M+H]^{+1}$ daughter ions of 1706.7, 1544.6, 1382.6, 1220.6, and 1058.5 *m/z* arising from neutral losses of 162 *m/z* (Fig. 5.3D). This provides strong evidence of a Chy_1 glycopeptide possessing five hexoses. The MS/MS spectra was noted as unusual in that a second ion

series with neutral losses of 162 *m/z* was also observed (1668.7, 1506.7, 1344.6, 1182.6, 1020.5, and 858.6 *m/z*), and the largest ion in this series (1668.7 *m/z*) differed from the parent ion by 201.1 *m/z*. This series corresponds to glycosylation of the y_8-ion (858.6 *m/z*) that resulted from fragmentation between $_{42}$Val-Pro$_{43}$.

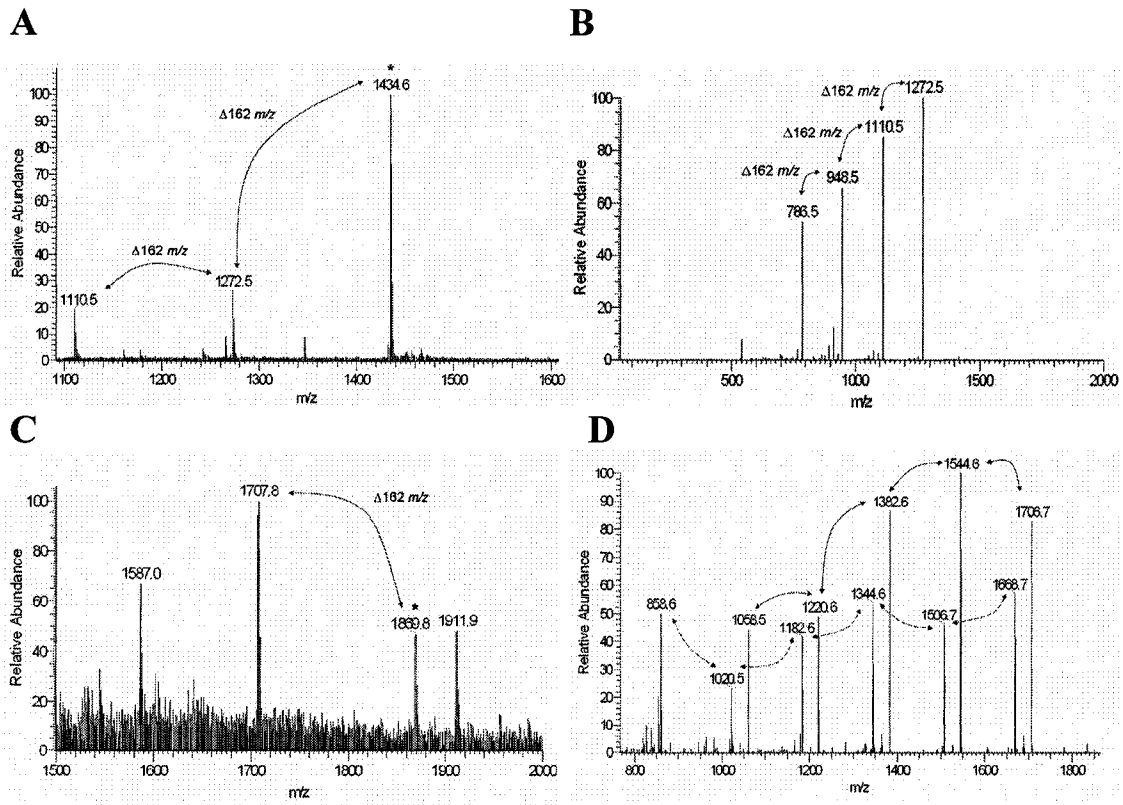


Fig. 5.3 MS and MS/MS of glycopeptides generated by chymotrypsin digestion of rSodC. (**A**) ESI-MS spectrum averaged over 80 scans corresponding to the Chy_2 ($_{51}$TGSPAPSGL$_{59}$) peptide demonstrates primarily glycosylated forms modified with two, three, and four hexoses. (**B**) The 1434.6 *m/z* $[M+H]^{+1}$ molecular ion marked by an asterisk in (A) was selected for ESI-MS/MS resulting in a fragmentation pattern containing neutral losses of 162 *m/z*. The 786.5 *m/z* molecular ion represents the fully deglycosylated peptide. (**C**) ESI-MS spectrum averaged over 96 scans corresponding to the Chy_1 ($_{41}$TVPGTTPSIW$_{50}$) peptide demonstrates primarily glycosylated forms modified with four and five hexoses. (**D**) The 1869.8 *m/z* $[M+H]^{+1}$ molecular ion marked by an asterisk in (C) was selected for ESI-MS/MS resulting in a fragmentation pattern containing neutral losses of 162 *m/z* (marked with solid arrows →). The 1058.5 *m/z* molecular ion represents the fully deglycosylated peptide. In addition, an ion series corresponding to a glycosylated y_8 ion series with differences of 162 *m/z* was observed (marked with dashed arrows --->).

The overall glycoprotein structure of rSodC from *Mtb* was supported by Matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) MS analysis. Specifically we observed singly-charged molecular ions of the glycosylated rSodC 28-kDa product (N-terminus of Thr_{41}) possessing between six to ten hexosyl residues (data not shown). Additionally, the mass spectrum shows a set of peaks assigned to singly-charged molecular ions of the nonglycosylated 23- and 22-kDa rSodC products (N-termini of Gly_{58} and Ser_{60}, respectively), and additional minor peaks assigned to nonglycosylated products with N-termini of Gly_{61}, Asp_{63}, Glu_{64}, Glu_{65}, Ser_{66}, and Gly_{68} (data not shown).

5.3.3 Expression of site-directed mutant rSodC proteins in *Mtb*

The MS-based analyses demonstrated SodC glycosylation localized to the N-terminus, and were in agreement of the differential ConA reactivity profiles and N-terminal sequencing of the 22-, 23-, and 28-kDa rSodC products from *Mtb.* These results are further supported by analysis of the SodC sequence with the NetOglyc 3.1 program, a neural network algorithm trained on eukaryotic mucin-type O-glycosylation sites (24) (Fig. 5.4). In general, the highest predictive scores for glycosylation were found in the N-terminal part of the protein. Specifically, four Thr and four Ser residues predicted to be O-glycosylated are upstream of the N-terminal amino acids of the non-ConA reactive 22- and 23-kDa products.

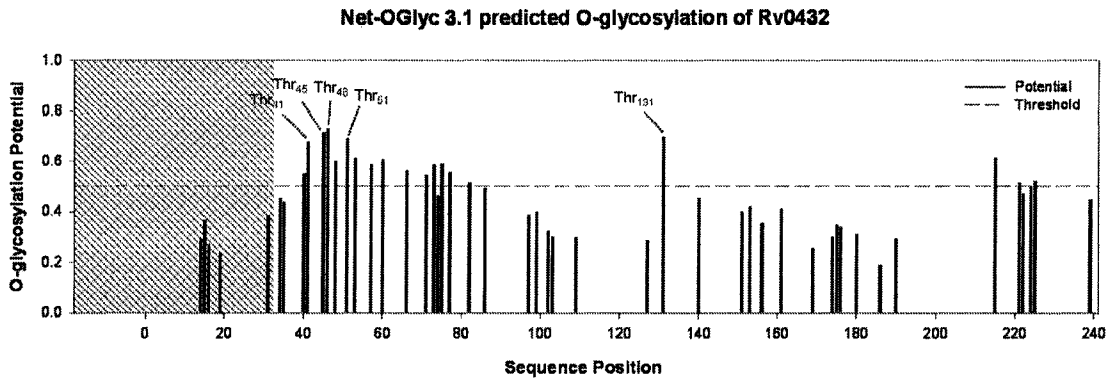
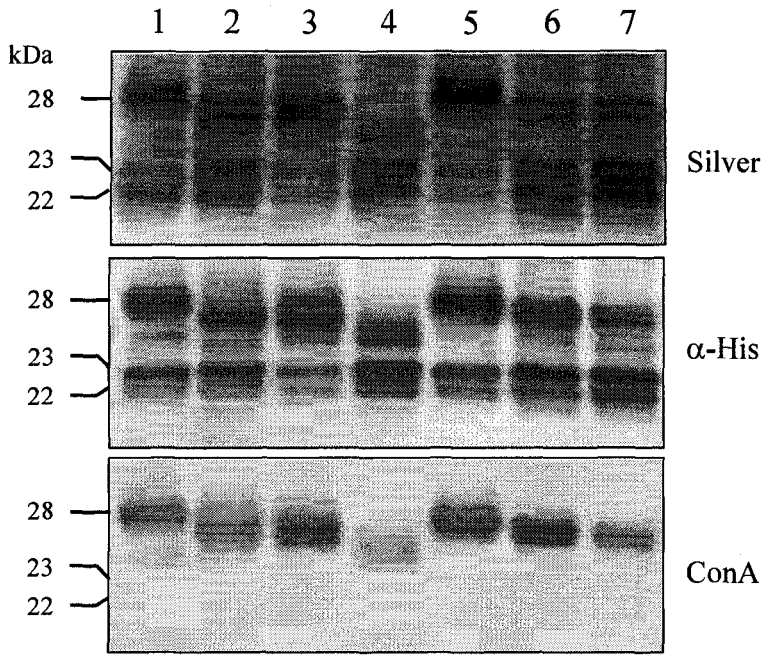


Fig. 5.4 Prediction of O-glycosylation sites for SodC. The complete Rv0432 protein sequence was analyzed using the NetOglyc 3.1 algorithm (24). The shaded region indicates the signal peptide. The labeled Thr residues indicate the predicted glycosylation sites targeted for amino acid substitution with Ala.

To identify specific O-glycosylated residues of SodC, those amino acids with the highest predictive NetOglyc scores (Thr_{41}, Thr_{45}, Thr_{46}, and Thr_{51}) were selected for site-directed mutagenesis to allow substitution of Thr with Ala. The more C-terminal Thr_{131} residue was also targeted as this residue had a high predictive score for glycosylation. Additionally, a double mutant in $_{45}ThrThr_{46}$ was constructed. Episomal copies of the *sodC* gene containing the site-directed mutations were expressed in *Mtb* and proteins were purified from the CF. Analysis of the recombinant proteins by SDS-PAGE, and Western blot revealed each protein was represented by multiple bands (Fig. 5.5A). Like the non-mutated rSodC, ConA-negative products of 22- and 23-kDa were observed for each of the rSodC products possessing Thr to alanine (Ala) substitutions. The relative abundance of these two products varied among the protein constructs, but no differences in electrophoretic migration were observed. Variations in molecular mass, however, were observed when comparing the non-mutated 28-kDa rSodC product to the corresponding products with Thr to Ala substitutions (Fig. 5.5A). Specifically, all of the

altered products except rSodC-Thr_{131}-Ala_{131} displayed some reduction in mass. The differences in mass did not appear to be due to truncation at the C-terminus since all of the higher molecular mass products displayed similar reactivity against the anti-His_5 monoclonal antibody. To ascertain differences in the N-termini, each high-mass product (25-28 kDa) was subjected to N-terminal sequencing (Fig. 5.5B). Four of the six mutated proteins (rSodC-Thr_{45}-Ala_{45}, rSodC-Thr_{46}-Ala_{46}, rSodC-Thr_{51}-Ala_{51}, and rSodC-Thr_{131}-Ala_{131}) possessed N-terminal sequences identical to the non-mutated rSodC, and rSodC-Thr_{41}-Ala_{41} differed only in the loss of the N-terminal Thr_{41}. In contrast, a dramatic difference was observed in the N-terminal sequence of rSodC-$Thr_{45}Thr_{46}$-$Ala_{45}Ala_{46}$ product where a four amino acid truncation occurred as compared to non-mutated rSodC, and SDS-PAGE migration indicated a loss of approximately 2 kDa in mass (Fig. 5.5A, lane 4 and Fig. 5.5B). Recombinant rSodC-$Thr_{45}Thr_{46}$-$Ala_{45}Ala_{46}$ was also the only protein construct that showed a significant decrease in ConA reactivity. Therefore, the observed reduction in the mass of rSodC-$Thr_{45}Thr_{46}$-$Ala_{45}Ala_{46}$ was attributed to the absence of $_{41}TVPG_{44}$ and decreased glycosylation. A qualitative decrease in ConA reactivity was not as apparent for the other mutated proteins rSodC-Thr_{45}-Ala_{45}, rSodC-Thr_{46}-Ala_{46}, rSodC-Thr_{51}-Ala_{51}, and rSodC-Thr_{41}-Ala_{41}) that had modest reductions in their molecular masses. Nevertheless, decreased glycosylation was speculated to be the cause for these molecular mass shifts since rSodC-Thr45Thr46-Ala45Ala46 retained modest levels of ConA reactivity. It was noted that amino acid assignment by Edman degradation sequencing was consistently difficult for the positions corresponding to Thr_{45}, Thr_{46}, and Ser_{48}, a known characteristic of O-glycosylated residues (1).

A



B

	Reference sequence	C - S - S - P - Q - H - A - S - T^{41} - V - P - G - T^{45} - T^{46} - P - S - I - W - T^{51} - G - S
Lane	Product	
1	rSodC	T - V - P - G - X - X - P - X - I
2	rSodC-Thr_{45}-Ala_{45}	T - V - P - G - A
3	rSodC-Thr_{46}-Ala_{46}	T - V - P - G - X
4	rSodC-$Thr_{45}Thr_{46}$-$Ala_{45}Ala_{46}$	A - A - P - X
5	rSodC-Thr_{131}-Ala_{131}	T - V - P - G - X
6	rSodC-Thr_{41}-Ala_{41}	V - P - G - X - X
7	rSodC-Thr_{51}-Ala_{51}	T - V - P - G - X

Fig. 5.5 Analyses of purified rSodC proteins possessing Thr to Ala substitutions. (**A**) Silver-stained SDS-PAGE (top panel), and Western blot analyses with anti-His_5 monoclonal antibody (middle panel) and ConA (bottom panel) as probes. Lane 1, non-mutated rSodC; lane 2, rSodC-Thr_{45}-Ala_{45}; lane 3, rSodC-Thr_{46}-Ala_{46}; Lane 4, rSodC-$Thr_{45}Thr_{46}$-$Ala_{45}Ala_{46}$; Lane 5, rSodC-Thr_{131}-Ala_{131}; Lane 6, rSodC-Thr_{41}-Ala_{41}; and Lane 7, rSodC-Thr_{51}-Ala_{51}. (**B**) Amino-terminal sequencing of purified rSodC products. An "X" denotes the inability to assign an amino acid. The reference sequence provided begins with the putative lipidated Cys_{33} residue following signal peptide cleavage. Only the high-mass (25- to 28-kDa) products were sequenced.

5.3.4 MS-based analyses of mutant rSodC proteins

Loss in ConA reactivity and molecular mass are potential indicators of decreased protein glycosylation; however, MS analyses provide a more definitive measure of the level of glycosylation as well as the ability to identify specific O-glycosylated residues. Thus, chymotrypsin digestions of the mutant rSodC proteins were analyzed by LC-ESI-MS/MS and the data was analyzed in the same manner as that obtained for the glycopeptides of the non-mutated rSodC.

Searching the rSodC-Thr_{51}-Ala_{51} MS/MS data for spectra arising from neutral losses of 162 *m/z* again identified glycosylated Chy_2 peptides. Three major $[M+H]^{+1}$ molecular ions (918.5, 1080.6, and 1242.5 *m/z*) that differed by 162 *m/z* were found for the $_{51}$AGSPAPSGL$_{59}$ peptide, corresponding to modifications of one, two, and three hexoses, respectively (Fig. 5.6A). Fragmentation and MS/MS of the largest molecular ion (1242.5 *m/z*) generated the $[M+H]^{+1}$ daughter ions of 1080.5, 918.5, and 756.5 *m/z* arising from neutral losses of 162 *m/z* (Fig. 5.6B). Therefore, the Thr_{51} to Ala_{51} substitution resulted in the loss of only one hexose compared to non-mutated rSodC, indicating that Thr_{51} possessed a single hexose and either or both Ser_{53} and Ser_{57} were modified with one to three hexoses, an unexpected finding. This was supported by the MS/MS data of the non-mutated rSodC Chy_2 peptide where the y_4, y_6, and y_7 fragmentation ions corresponding to glycosylated forms of $_{56}$PSGL$_{59}$, $_{54}$PAPSGL$_{59}$, and $_{53}$SPAPSGL$_{59}$, respectively were observed (data not shown).

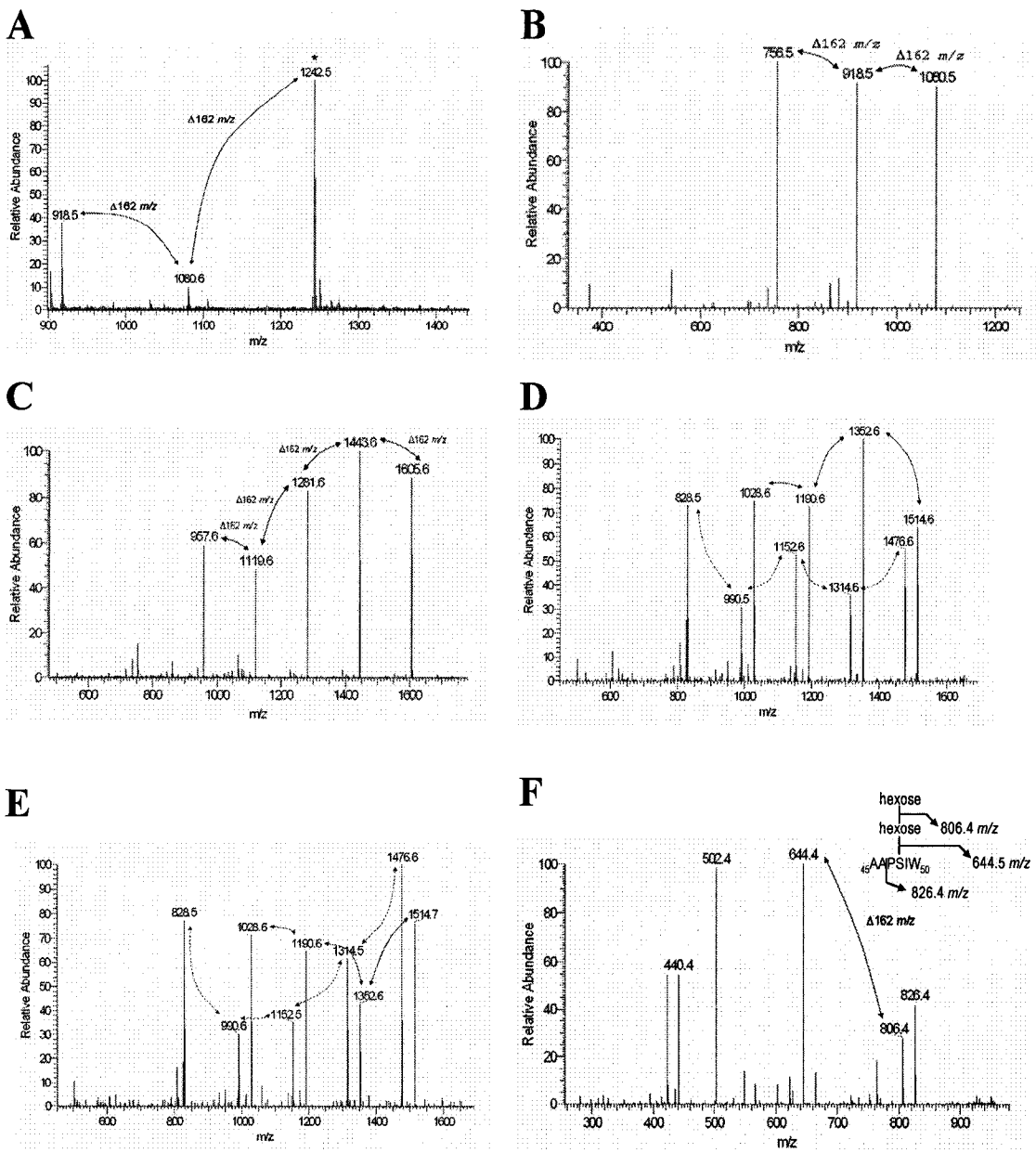


Fig. 5.6 MS and MS/MS of glycopeptides generated by chymotrypsin digestion of rSodC products with Thr to Ala substitutions. (**A**) MS-ESI spectrum averaged over 89 scans corresponding to Chy_2 ($_{51}$AGSPAPSGL$_{59}$) peptide of rSodC-Thr$_{51}$-Ala$_{51}$ demonstrates primarily glycosylated forms modified with one to three hexoses. (**B**) The 1242.5 *m/z* $[M+H]^{+1}$ molecular ion marked by an asterisk in (A) was selected for ESI-MS/MS resulting in a fragmentation pattern containing neutral losses of 162 *m/z*. The 756.5 *m/z* molecular ion represents the fully deglycosylated peptide. (**C**) The 1768.8 *m/z* $[M+H]^{+1}$ molecular ion corresponding to the Chy_1 ($_{42}$VPGTTPSIW$_{50}$) peptide of rSodC-Thr$_{41}$-Ala$_{41}$ with five hexoses was selected for ESI-MS/MS resulting in a fragmentation pattern containing neutral losses of 162 *m/z*. The 957.6 *m/z* molecular ion represents the fully deglycosylated peptide. (**D**) The 1677.7 *m/z* $[M+H]^{+1}$ molecular ion corresponding

to the Chy_1 ($_{41}$TVPGATPSIW$_{50}$) peptide of rSodC-Thr_{45}-Ala_{45} with four hexoses was selected for ESI-MS/MS resulting in a fragmentation pattern containing neutral losses of 162 *m/z* (marked with solid arrows →). The 1028.6 *m/z* molecular ion represents the fully deglycosylated peptide. In addition, an ion series corresponding to a glycosylated y_8 ion series with differences of 162 *m/z* was observed (marked with dashed arrows --->). (**E**) The 1677.7 *m/z* $[M+H]^{+1}$ molecular ion corresponding to the Chy_1 ($_{41}$TVPGTAPSIW$_{50}$) peptide of rSodC-Thr_{46}-Ala_{46} with four hexoses was selected for ESI-MS/MS resulting in a fragmentation pattern containing neutral losses of 162 *m/z* (marked with solid arrows →). The 1028.6 *m/z* molecular ion represents the fully deglycosylated peptide. In addition, an ion series corresponding to a glycosylated y_8 ion series with differences of 162 *m/z* was observed (marked with dashed arrows --->). (**F**) The 969.0 *m/z* $[M+H]^{+1}$ molecular ion corresponding to the Chy_1 ($_{45}$AAPSIW$_{50}$) peptide of rSodC-$Thr_{45}Thr_{46}$-$Ala_{45}Ala_{46}$ with two hexoses was selected for ESI-MS/MS resulting in a fragmentation pattern containing neutral losses of 162 *m/z*. The 644.4 *m/z* molecular ion represents the fully deglycosylated peptide. The inset shows the fragmentation of this glycopeptide.

Similar to non-mutated rSodC, the mutant rSodC Chy_1 peptides were consistently found in low abundance, and an approach was taken to selectively scan for $[M+H]^{+1}$ molecular ions corresponding to the predicted Chy_1 peptides modified with zero to five hexoses taking into account the experimentally determined N-termini. For rSodC-Thr_{41}-Ala_{41} two major $[M+H]^{+1}$ molecular ions (1606.7 and 1768.8 *m/z*) were found that corresponded to the $_{42}$VPGTTPSIW$_{50}$ peptide with four or five hexosyl residues, respectively. Fragmentation and MS/MS of the 1768.8 *m/z* ion generated the $[M+H]^{+1}$ daughter ions of 1605.6, 1443.6, 1281.6, 1119.6, and 957.6 *m/z* arising from neutral losses of 162 *m/z* (Fig. 5.6C). Thus N-terminal peptides of rSodC-Thr_{41}-Ala_{41} and non-mutated rSodC were similar in terms of the level of glycosylation, each with four to five hexoses, demonstrating Thr_{41} is not glycosylated. In support of this, N-terminal sequencing was consistently successful in assigning a Thr residue to this position in the non-mutated 28-kDa rSodC, an assignment otherwise blocked if this residue was glycosylated. For both rSodC-Thr_{45}-Ala_{45} and rSodC-Thr_{46}-Ala_{46} three major $[M+H]^{+1}$

molecular ions (1353.5, 1515.6, and 1677.7 *m/z*) were found that differed by 162 *m/z* and corresponded to the $_{41}$TVPGATPSIW$_{50}$ and $_{41}$TVPGTAPSIW$_{50}$ peptides with two, three, and four hexoses, respectively. Fragmentation and MS/MS of the 1677.7 *m/z* ion generated the $[M+H]^{+1}$ daughter ions of 1514.6-1514.7, 1352.6, 1190.6, and 1028.6 *m/z* arising from neutral losses of 162 *m/z* (Fig. 5.6D and 5.6E). This demonstrated that an amino acid substitution at Thr$_{45}$ or Thr$_{46}$ resulted in the loss of one hexosyl residue. Similar to the non-mutated rSodC Chy$_1$ peptide, these MS/MS spectra contained a second ion series with neutral losses of 162 *m/z* (1476.6, 1314.5-1314.6, 1152.6-1152.6, 990.5-990.6, and 828.5 *m/z*), and this ion series corresponded to glycosylation of the y_8-ion (828.5 *m/z*) that resulted from fragmentation between $_{42}$Val-Pro$_{43}$. For rSodC-Thr$_{45}$Th$_{r46}$-Ala$_{45}$Ala$_{46}$ two major $[M+H]^{+1}$ molecular ions (806.9 and 969.0 *m/z*) were found that differed by 162 *m/z* and corresponded to the $_{45}$AAPSIW$_{50}$ peptide with one and two hexoses, respectively. Fragmentation and MS/MS of the 969.0 *m/z* ion generated $[M+H]^{+1}$ daughter ions of 806.4 and 644.4 *m/z* demonstrating the presence of two hexoses on this peptide (Fig. 5.6F). In summary, the data generated for all mutants in the $_{45}$ThrThr$_{46}$ cluster were in agreement, and revealed each Thr residue is modified with one to two hexoses and collectively the $_{45}$ThrThr$_{46}$ doublet is modified with a total of three hexoses. The remaining glycosylation on the $_{45}$AAPSIW$_{50}$ peptide is only explained if Ser$_{48}$ is modified with one to two hexoses. A summary of the MS-based analyses of the mutated rSodC Chy$_1$ and Chy$_2$ peptides is shown in Table 5.3.

Table 5.3 Summary of MS-based analyses of rSodC glycopeptides with Thr to Ala substitutions

Protein	Relevant Peptide	Peptide Sequence	# of hexosyl residues[1]	Δ from non-mutated
non-mutated rSodC	Chy_1	TVPGTTPSIW	4-5	---
rSodC-Thr_{41}-Ala_{41}	Chy_1	VPGTTPSIW	4-5	0
rSodC-Thr_{45}-Ala_{45}	Chy_1	TVPGATPSIW	2-4	1-2
rSodC-Thr_{46}-Ala_{46}	Chy_1	TVPGTAPSIW	2-4	1-2
rSodC-$Thr_{45}Thr_{46}$-$Ala_{45}Ala_{46}$	Chy_1	AAPSIW	1-2	3
non-mutated rSodC	Chy_2	TGSPAPSGL	2-4	---
rSodC-Thr_{51}-Ala_{51}	Chy_2	AGSPAPSGL	1-3	1

[1]Based on the molecular ions observed in the full MS spectra

5.3.5 Subcellular localization

The SodC of *Mtb* is annotated as a lipoprotein, and would be expected to be associated with the cell wall or membrane of *Mtb*. However, we initially identified this antigenic protein in the CF of *Mtb*. To assess the subcellular location of SodC, *Mtb* H37Rv cytosol, membrane, cell wall, and CF fractions were isolated (21), separated by SDS-PAGE, electroblotted, and probed with polyclonal antiserum raised against SodC (Fig. 5.7A). Native SodC was dominant in the membrane and cell wall fractions, with a 27-kDa product being the most abundant form. The protein was also present in the CF, but in much smaller quantities (Fig. 5.7B). In the CF a 22-kDa product was also observed. The 1.1-kDa C-terminal $_{241}$KLHHHHHH$_{248}$ sequence present only in recombinant forms likely accounts for the difference in size between native and recombinant products. The rSodC showed a subcellular distribution pattern similar to the native protein (data not shown), and the 28-kDa rSodC product isolated from the WCL had an N-terminal sequence of $_{41}$TVPG$_{44}$. Additionally, molecular masses of the

recombinant products did not differ between WCL and CF forms from the same strain, indicating the WCL forms of these products were truncated in the same manner (data not shown). The native SodC was believed to be acylated in part based on detergent phase partitioning experiments (8). Unexpectedly, we found that both the native 27-kDa SodC and 28-kDa rSodC products from both WCL and CF partitioned to the detergent phase, even though these sequences lack the putative acylated Cys_{33} (Fig. 5.7B). To confirm this finding the *rv0432* gene without the signal sequence was expressed in *Mycobacterium smegmatis* and analysis of the 25-kDa SP-rSodC product in the WCL demonstrated that SodC partitions to the detergent phase even in the absence of an acylated Cys residue. A hydropathy plot of SodC suggests the protein sequence contains extended regions of hydrophobicity, explaining this behavior (data not shown), and brings into question whether this protein is truly acylated.

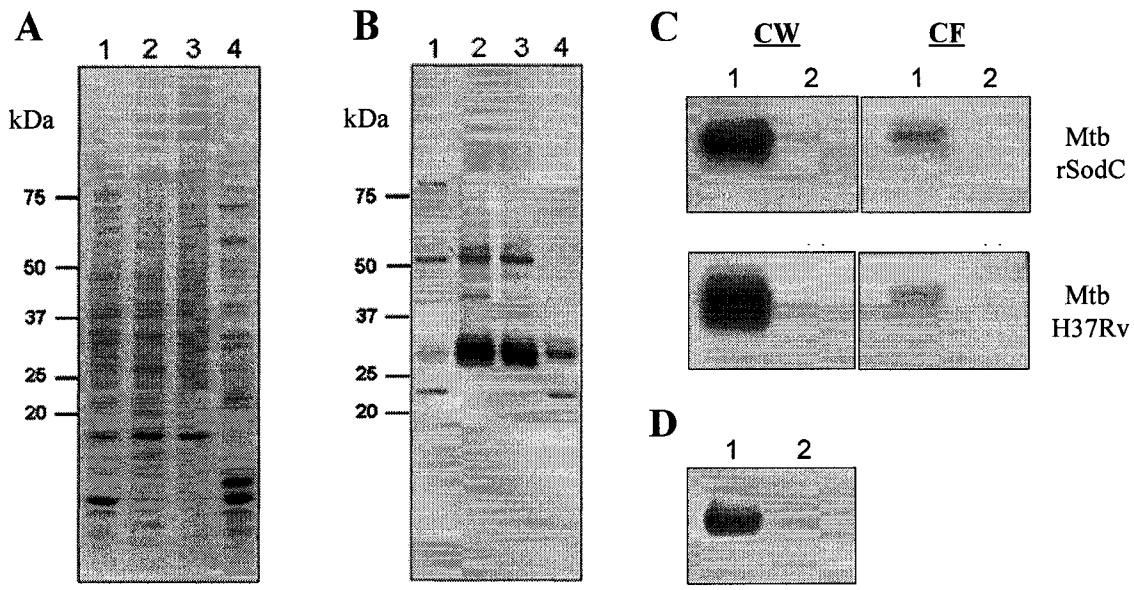


Fig. 5.7 SodC subcellular localization. Subcellular fractions of *Mtb* H37Rv (lane 1, cytosol; lane 2, membrane; lane 3, cell wall; and lane 4, CF) were separated by SDS-PAGE and stained with Coomassie Brilliant Blue (**A**) or electroblotted and probed with polyclonal SodC antiserum (**B**). (**C**) Western blot with polyclonal SodC antiserum of TX-114 detergent phase partitioning (lane 1, detergent phase; lane 2, aqueous phase) performed on cell wall (left panels) and CF (right panels) fractions of *Mtb* expressing rSodC (top panels) and wild type *Mtb* H37Rv (bottom panels). (**D**) Western blot with polyclonal SodC antiserum of TX-114 detergent phase partitioning (lane 1, detergent phase; lane 2, aqueous phase) performed on WCL of *M. smegmatis* SP-rSodC. *CW*, cell wall

5.4 Discussion

The detailed evaluation and description of glycosylation patterns for the MPT32 and MPB83 lipoprotein was accomplished with native products from filtrates of *in-vitro*-grown cells (11, 12, 30). In contrast SodC is not a dominant protein of the CF, and previously escaped detection in proteome-wide evaluation of *Mtb* CF proteins (25, 39, 45), and was only found in this extracellular preparation after extensive enrichment via multi-dimensional chromatography (Chapter 3). Thus, to obtain sufficient quantities for analyses the *sodC* gene was over expressed in its natural host (*Mtb*) and recombinant

protein purified from the CF. N-terminal sequencing and analysis of these products with the lectin ConA suggested that glycosylation was localized to the 19-amino-acid N-terminus, and the NetOglyc algorithm (24) predicted specific Ser and Thr residues to be O-glycosylated in this region. Taking these observations into account and considering the previous discovery of only O-glycosylated Thr residues in mycobacteria (11, 12, 30), five Thr residues of SodC with the highest predictive scores were substituted with Ala. Although Val was selected for substitutions of putative glycosylated Thr residues in previous studies (20, 30), we found that Ala substitutions had a more neutral effect in terms of influencing the glycosylation of neighboring residues as predicted by the NetOglyc algorithm (data not shown), and also resulted in a mass difference (30 Da) easily tracked in our MS experiments. None of the Thr to Ala substitutions resulted in a complete loss of glycosylation as assessed by ConA reactivity and MS. Nevertheless, the substitutions allowed for the assignment of glycosylated residues and determination of the extent of glycosylation at each residue. Of particular interest was the identification of O-glycosylated Ser residues which is in contrast to earlier studies where only Thr residues of *Mtb* proteins were shown to be glycosylated (11, 12, 20, 30). Arguably the Thr to Ala substitutions might have altered the natural glycosylation state of surrounding sequences and in effect forced glycosylation onto neighboring Ser residues. However, MS/MS data of the non-mutated rSodC Chy_2 peptide showed this phenomenon was not occurring. Our data located three Thr and two to three Ser residues possessing modifications with one to three hexose units, with the total number of hexose units ranging from six to ten. Figure 5.8 summarizes these results.

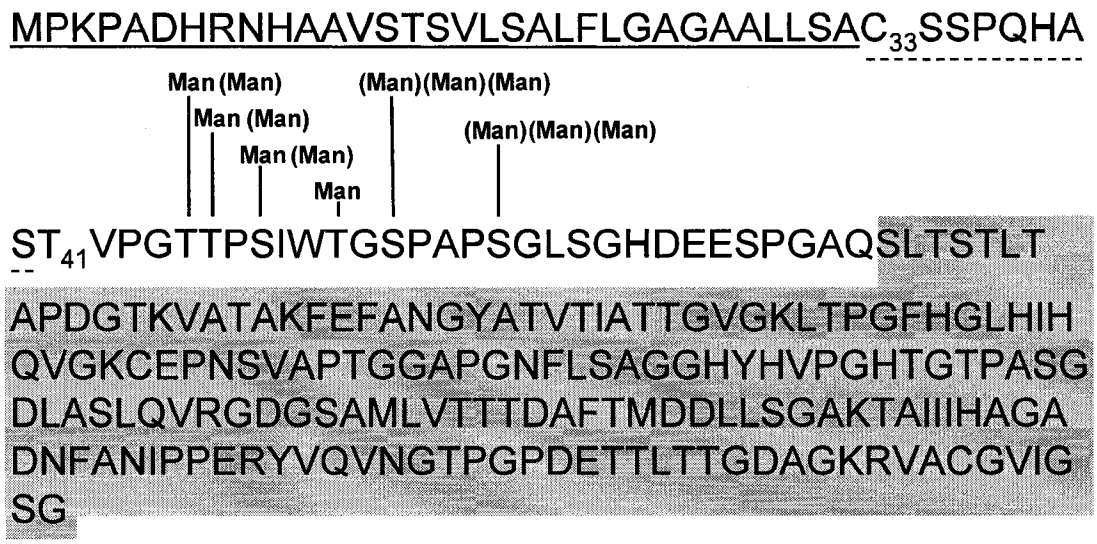


Fig. 5.8 Working model for the posttranslational modification of *Mtb* SodC. The region underlined with a solid line denotes the Type II signal peptide preceding the putative N-terminal acylated Cys (C_{33}) described in ref (8). The region underlined with the dashed line denotes the experimentally determined truncation that resulted in an N-terminal Thr (T_{41}). Glycosylation sites modified with mannose are indicated. Parentheses denote variable levels of glycosylation. The shaded region indicates the folded enzymatic structure, for which a more detailed annotation can be found in ref (46). *Man*, mannose

The glycosylation pattern of SodC showed several similarities to previously described mycobacterial glycoproteins. Consistent with oligosaccharide lengths of MPT32 and MPB83, each site of O-glycosylation on SodC was populated with a mono-, di- or tri-hexose. Although we did not provide direct chemical evidence for the identity of the modifying sugar(s) on SodC, mannose was likely the glycosylating entity given reactivity to ConA, the observed neutral losses of 162 *m/z,* and the specificity of the jack bean α-mannosidase used to deglycosylate the protein. Similar to the 19-kDa lipoprotein, MPT32, and MPB83, a Thr doublet near the N-terminus was glycosylated (12, 20, 30). Furthermore, the SodC Thr doublet showed great similarity to the MPB83 Thr doublet, where each Thr is variably modified with a mannose or a mannobiose, and

the Thr doublet is collectively modified with a total of three mannose residues (30). Features of the five SodC glycosylation sites (treating the Thr doublet as one site) were similar to the MPB83 and MPT32 glycosylation-site structural features described by (4). Specifically, three of the five glycosylation sites were present one or two amino acids downstream of a Pro residue, and all five sites were localized to Pro, Ala, Thr, and Ser rich domains. Finally, similar to the other glycoproteins of *Mtb*, SodC possesses multiple O-glycosylated residues within a short region (five to six residues in a 13 amino-acid sequence).

SodC is considered one of the few well characterized lipoproteins of *Mtb* (47) and evidence for its acylation comes from radiolabeling experiments on recombinant SodC produced in *E. coli* and a subsequent comparisons of recombinant and native protein by non-denaturing activity gel electrophoresis (8). It should also be noted that in the single study to assess native *Mtb* lipoproteins by metabolic labeling, products of 26 and 27 kDa were observed (54). We found native SodC to be equally distributed between the membrane and cell wall fractions, and this was likely due to mixing of the membrane and cell wall subcellular fractions during the mechanical breakage of cells. The association of SodC with the cell envelope is in agreement with the study of Wu *et al.* (53). However, we unexpectedly found cell-associated rSodC to possess the same N-terminal sequence, 2-D gel electrophoresis pattern, and MS-based glycosylation pattern as the non-lipidated rSodC purified from the CF (data not shown). The described alterations in SDS-PAGE migration resulting from each Thr to Ala substitution were also observed for the cell-associated- as well as the CF- form of rSodC. These data (in particular, that from N-terminal sequencing) demonstrated that cell-associated rSodC was not lipidated. The

association of SodC with the cell wall in the absence of acylation may result from interactions between extended hydrophobic stretches of the protein with membrane lipids. This possibility is supported by our observations that the non-acylated rSodC localized to the detergent fraction of a TX-114 biphasic partitioning. Alternatively, the cell wall association could result from hydrophilic interactions of the glycosylated N-terminus with carbohydrate constituents of the cell wall. Such a phenomenon would also explain the recent observation that that the non-lipidated glycoprotein MPT32/Apa associates with the cell wall (37).

Two hypotheses can account for the presence of non-lipidated SodC products. The first involves multiple signal peptidases acting on the SodC signal peptide. As a lipoprotein, SodC is predicted to be processed by LspA/Rv1539, a lipoprotein-specific SPase II (41). Processing of SodC by the *Mtb* type-I signal peptidase (SPase I/LepB/Rv2903c) could also occur. Evaluation of the SodC sequence by SignalP (34) reveals a potential cleavage site between Ala_{39} and Ser_{40} as well as at the putative acylated Cys_{33}. (data not shown). The alternative processing by SPase I may be most relevant to the N-terminal sequence data obtained for the rSodC-Thr_{41}-Ala_{41} construct since the substitution made led to a predicted SPase I cleavage site between Thr_{41} and Val_{42} that corresponds to the experimentally determined N-terminus of this protein. Another possibility is that SPase II may have alternative cleavage sites, or acts prior to acylation as has been recently described for *Listeria monocytogenes* (3).

The second hypothesis that could account for the lack of acylated rSodC is that of "proteolytic shaving" (50). This has been reported for *Bacillus subtilis* where amino acids downstream of the acylated N-terminal Cys direct the release of lipoproteins from

membranes through proteolytic cleavage. However, in the case of the *B. subtilis* protein glycosylation was not involved. Other have proposed that glycosylation of *Mtb* proteins may serve as a signal for proteolytic cleavage, or that glycosylation prevents further amino-terminal degradation following proteolytic cleavage from the acylated anchor (20, 30). Interestingly, similar to the 25-kDa native CF form of MPB83, the larger rSodC in the CF has an N-terminal Thr preceding the glycosylated Thr doublet ($Thr_{45}Thr_{46}$) and substitution of these Thr residues to Ala resulted in a new N-terminus and further truncation of SodC. This also was similar to the truncation reported for the 19-kDa lipoprotein when two glycosylated Thr clusters were substituted with Val residues (20). Thus, with the data for the rSodC protein produced in *Mtb* there is further evidence that amino acids in the N-terminal region and downstream of the acylated Cys influence "proteolytic shaving" and that glycosylation further influences this process.

A large number of SodC homologues from multiple prokaryotic and eukaryotic species have been studied and the N-terminal glycosylated region of *Mtb* SodC is conserved among SodC homologues of actinomycetes (Fig. 5.9A), but not with Cu,Zn SODs of other species (Fig. 5.9B). Further, the *Mtb* SodC is the only member of the Cu,Zn SOD family shown to be glycosylated. This would suggest that the N-terminal portion of the mycobacterial SodC proteins serves a unique function. Recent elucidation of the crystal structure of a secreted non-acylated form of *Mtb* SodC produced in *E. coli* (46) failed to provide a defined structure of the N-terminal region (Cys_{33} and Gln_{70}) of the mature protein and this was attributed to the unordered and flexible characteristics of this region. Prediction of intrinsically folded sequences within the SodC mature protein sequence using the FoldIndex© tool (36) agreed with this observation (data not shown).

Thus, there is likely a physical separation between the unordered glycosylated N-terminus of SodC and its folded enzymatic domain; suggesting that glycosylation does not directly influence the catalytic activity of SodC. In fact a recombinant form of this protein produced in *E. coli* where glycosylation was not possible retained catalytic activity based on a qualitative assay (8). O-glycosylation is proposed to contribute to an unordered and extended conformation in proteins (23). Thus, the predicted structure of the O-glycosylated region of SodC lends support to the above hypothesis that glycosylation modulates proteolytic processing of the unstructured SodC N-terminus and greatly influences SodC localization.

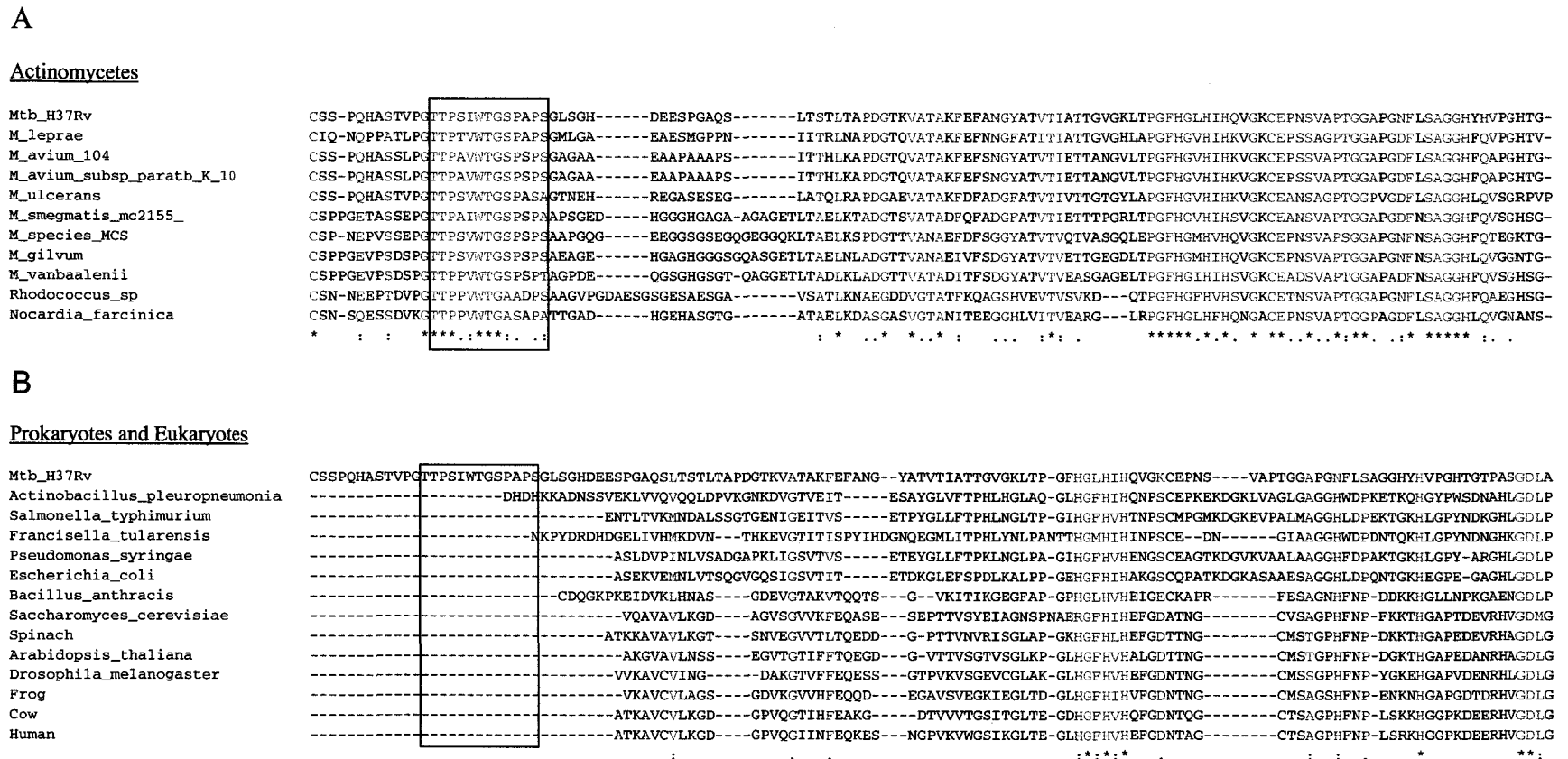


Fig. 5.9 Partial SodC protein sequence alignment from select species. Two sequence-based alignments of homologous SodC sequences from (**A**) actinomycetes and (**B**) prokaryotic and eukaryotic species were generated with the ClustalW software (49). Shown are sequences corresponding to Cys_{33} to Ala_{160} from *Mtb* SodC. For prokaryotic species, the signal peptides were removed prior to alignment. Identical (*), strongly similar (:), and weakly similar (.) residues are indicated. The boxed region corresponds to the *Mtb* SodC glycosylated sequence. Modified from Spagnolo *et al.* (46).

5.5 Literature cited

1. **Abernathy, J. L., Wang, Y., Eckhardt, A.E., and Hill, R.L.**, Identification of O-glycosylation sites with a gas phase sequencer, in Techniques in Protein Chemistry, J.L. Abernathy, Editor. 1992, Academic Press, Inc.: New York. p. 277-86.

2. **Battistoni, A. and G. Rotilio**. 1995. Isolation of an active and heat-stable monomeric form of Cu,Zn superoxide dismutase from the periplasmic space of Escherichia coli. FEBS Lett **374**(2): 199-202.

3. **Baumgartner, M., U. Karst, B. Gerstel, M. Loessner, J. Wehland, and L. Jansch**. 2007. Inactivation of Lgt allows systematic characterization of lipoproteins from Listeria monocytogenes. J Bacteriol **189**(2): 313-24.

4. **Belisle, J. T., M. Braunstein, I. Rosenkrands, and P. Andersen**, The Proteome of *Mycobacterium tuberculosis*, in Tuberculosis and the Tubercle Bacillus, S.T. Cole, Editor. 2005, ASM Press: Washingoton, DC. p. 235-60.

5. **Braunstein, M., B. J. Espinosa, J. Chan, J. T. Belisle, and W. R. Jacobs, Jr.** 2003. SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. Mol Microbiol **48**(2): 453-64.

6. **Coligan, J. E. e. a.**, in Current Protocols in Protein Science. p. 10.5.1.

7. **Cutalo, J. M., L. J. Deterding, and K. B. Tomer**. 2004. Characterization of glycopeptides from HIV-I(SF2) gp120 by liquid chromatography mass spectrometry. J Am Soc Mass Spectrom **15**(11): 1545-55.

8. **D'Orazio, M., S. Folcarelli, F. Mariani, V. Colizzi, G. Rotilio, and A. Battistoni**. 2001. Lipid modification of the Cu,Zn superoxide dismutase from *Mycobacterium tuberculosis*. Biochem J **359**(Pt 1): 17-22.

9. **Daugelat, S., J. Kowall, J. Mattow, D. Bumann, R. Winter, R. Hurwitz, and S. H. Kaufmann**. 2003. The RD1 proteins of *Mycobacterium tuberculosis*: expression in *Mycobacterium smegmatis* and biochemical characterization. Microbes Infect **5**(12): 1082-95.

10. **Deshpande, R. G., M. B. Khan, D. A. Bhat, and R. G. Navalkar**. 1993. Superoxide dismutase activity of *Mycobacterium tuberculosis* isolated from tuberculosis patients and the immunoreactivity of superoxide dismutase from M. tuberculosis H37Rv. Tuber Lung Dis **74**(6): 388-94.

11. **Dobos, K. M., K. H. Khoo, K. M. Swiderek, P. J. Brennan, and J. T. Belisle**. 1996. Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. J Bacteriol **178**(9): 2498-506.

12. **Dobos, K. M., K. Swiderek, K. H. Khoo, P. J. Brennan, and J. T. Belisle**. 1995. Evidence for glycosylation sites on the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. Infect Immun **63**(8): 2846-53.

13. **Edwards, K. M., M. H. Cynamon, R. K. Voladri, C. C. Hager, M. S. DeStefano, K. T. Tham, D. L. Lakey, M. R. Bochan, and D. S. Kernodle**. 2001. Iron-cofactored superoxide dismutase inhibits host responses to *Mycobacterium tuberculosis*. Am J Respir Crit Care Med **164**(12): 2213-9.

14. **Espitia, C. and R. Mancilla**. 1989. Identification, isolation and partial characterization of *Mycobacterium tuberculosis* glycoprotein antigens. Clin Exp Immunol **77**(3): 378-83.

15. **Fenyo, D. and R. C. Beavis**. 2003. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Anal Chem **75**(4): 768-74.

16. **Fifis, T., C. Costopoulos, A. J. Radford, A. Bacic, and P. R. Wood**. 1991. Purification and characterization of major antigens from a *Mycobacterium bovis* culture filtrate. Infect Immun **59**(3): 800-7.

17. **Garbe, T., D. Harris, M. Vordermeier, R. Lathigra, J. Ivanyi, and D. Young**. 1993. Expression of the *Mycobacterium tuberculosis* 19-kilodalton antigen in *Mycobacterium smegmatis*: immunological analysis and evidence of glycosylation. Infect Immun **61**(1): 260-7.

18. **Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.**, Protein Identification and Analysis Tools on the ExPASy Server, in The Proteomics Protocols Handbook, J.M. Walker, Editor. 2005, Humana Press. p. 571-607.

19. **Herrmann, J. L., R. Delahay, A. Gallagher, B. Robertson, and D. Young**. 2000. Analysis of post-translational modification of mycobacterial proteins using a cassette expression system. FEBS Lett **473**(3): 358-62.

20. **Herrmann, J. L., P. O'Gaora, A. Gallagher, J. E. Thole, and D. B. Young**. 1996. Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from *Mycobacterium tuberculosis*. Embo J **15**(14): 3547-54.

21. **Hirschfield, G. R., M. McNeil, and P. J. Brennan**. 1990. Peptidoglycan-associated polypeptides of *Mycobacterium tuberculosis*. J Bacteriol **172**(2): 1005-13.

22. **Horn, C., A. Namane, P. Pescher, M. Riviere, F. Romain, G. Puzo, O. Barzu, and G. Marchal**. 1999. Decreased capacity of recombinant 45/47-kDa molecules (Apa) of *Mycobacterium tuberculosis* to stimulate T lymphocyte responses related to changes in their mannosylation pattern. J Biol Chem **274**(45): 32023-30.

23. **Jentoft, N.** 1990. Why are proteins O-glycosylated? Trends Biochem Sci **15**(8): 291-4.

24. **Julenius, K., A. Molgaard, R. Gupta, and S. Brunak**. 2005. Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology **15**(2): 153-64.

25. **Jungblut, P. R., U. E. Schaible, H. J. Mollenkopf, U. Zimny-Arndt, B. Raupach, J. Mattow, P. Halada, S. Lamer, K. Hagens, and S. H. Kaufmann**. 1999. Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. Mol Microbiol **33**(6): 1103-17.

26. **Kimura, Y., D. Hess, and A. Sturm**. 1999. The N-glycans of jack bean alpha-mannosidase. Structure, topology and function. Eur J Biochem **264**(1): 168-75.

27. **Kyte, J. and R. F. Doolittle**. 1982. A simple method for displaying the hydropathic character of a protein. J Mol Biol **157**(1): 105-32.

28. **Link, A. J., K. Robison, and G. M. Church**. 1997. Comparing the predicted and observed properties of proteins encoded in the genome of Escherichia coli K-12. Electrophoresis **18**(8): 1259-313.

29. **McDonald, W. H. and J. R. Yates, 3rd**. 2002. Shotgun proteomics and biomarker discovery. Dis Markers **18**(2): 99-105.

30. **Michell, S. L., A. O. Whelan, P. R. Wheeler, M. Panico, R. L. Easton, A. T. Etienne, S. M. Haslam, A. Dell, H. R. Morris, A. J. Reason, J. L. Herrmann, D. B. Young, and R. G. Hewinson**. 2003. The MPB83 antigen from *Mycobacterium bovis* contains O-linked mannose and (1-->3)-mannobiose moieties. J Biol Chem **278**(18): 16423-32.

31. **Morrissey, J. H.** 1981. Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. Anal Biochem **117**(2): 307-10.

32. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne**. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng **10**(1): 1-6.

33. **Nielsen, H. and A. Krogh**. 1998. Prediction of signal peptides and signal anchors by a hidden Markov model. Proc Int Conf Intell Syst Mol Biol **6**: 122-30.

34. **Nielsen, J. B. and J. O. Lampen**. 1982. Glyceride-cysteine lipoproteins and secretion by Gram-positive bacteria. J Bacteriol **152**(1): 315-22.

35. **Piddington, D. L., F. C. Fang, T. Laessig, A. M. Cooper, I. M. Orme, and N. A. Buchmeier**. 2001. Cu,Zn superoxide dismutase of *Mycobacterium*

tuberculosis contributes to survival in activated macrophages that are generating an oxidative burst. Infect Immun **69**(8): 4980-7.

36. **Prilusky, J., C. E. Felder, T. Zeev-Ben-Mordehai, E. H. Rydberg, O. Man, J. S. Beckmann, I. Silman, and J. L. Sussman**. 2005. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics **21**(16): 3435-8.

37. **Ragas, A., L. Roussel, G. Puzo, and M. Riviere**. 2007. The *Mycobacterium tuberculosis* cell-surface glycoprotein apa as a potential adhesin to colonize target cells via the innate immune system pulmonary C-type lectin surfactant protein A. J Biol Chem **282**(8): 5133-42.

38. **Romain, F., C. Horn, P. Pescher, A. Namane, M. Riviere, G. Puzo, O. Barzu, and G. Marchal**. 1999. Deglycosylation of the 45/47-kilodalton antigen complex of *Mycobacterium tuberculosis* decreases its capacity to elicit in vivo or in vitro cellular immune responses. Infect Immun **67**(11): 5567-72.

39. **Rosenkrands, I., A. King, K. Weldingh, M. Moniatte, E. Moertz, and P. Andersen**. 2000. Towards the proteome of *Mycobacterium tuberculosis*. Electrophoresis **21**(17): 3740-56.

40. **Sambrook J.E., F. E. F., Maniatis T.**, Molecular Cloning: A Laboratory Manual. 1989, New York: Cold Spring Harbor Laboratory Press.

41. **Sander, P., M. Rezwan, B. Walker, S. K. Rampini, R. M. Kroppenstedt, S. Ehlers, C. Keller, J. R. Keeble, M. Hagemeier, M. J. Colston, B. Springer, and E. C. Bottger**. 2004. Lipoprotein processing is required for virulence of *Mycobacterium tuberculosis*. Mol Microbiol **52**(6): 1543-52.

42. **Shively, J. E., P. Miller, and M. Ronk**. 1987. Microsequence analysis of peptides and proteins. VI. A continuous flow reactor for sample concentration and sequence analysis. Anal Biochem **163**(2): 517-29.

43. **Smith, P. K., R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson, and D. C. Klenk**. 1985. Measurement of protein using bicinchoninic acid. Anal Biochem **150**(1): 76-85.

44. **Snapper, S. B., R. E. Melton, S. Mustafa, T. Kieser, and W. R. Jacobs, Jr.** 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. Mol Microbiol **4**(11): 1911-9.

45. **Sonnenberg, M. G. and J. T. Belisle**. 1997. Definition of *Mycobacterium tuberculosis* culture filtrate proteins by two-dimensional polyacrylamide gel electrophoresis, N-terminal amino acid sequencing, and electrospray mass spectrometry. Infect Immun **65**(11): 4515-24.

46. **Spagnolo, L., I. Toro, M. D'Orazio, P. O'Neill, J. Z. Pedersen, O. Carugo, G. Rotilio, A. Battistoni, and K. Djinovic-Carugo**. 2004. Unique features of the sodC-encoded superoxide dismutase from *Mycobacterium tuberculosis*, a fully functional copper-containing enzyme lacking zinc in the active site. J Biol Chem **279**(32): 33447-55.

47. **Sutcliffe, I. C. and D. J. Harrington**. 2004. Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. FEMS Microbiol Rev **28**(5): 645-59.

48. **Takayama, K., H. K. Schnoes, E. L. Armstrong, and R. W. Boyle**. 1975. Site of inhibitory action of isoniazid in the synthesis of mycolic acids in *Mycobacterium tuberculosis*. J Lipid Res **16**(4): 308-17.

49. **Thompson, J. D., D. G. Higgins, and T. J. Gibson**. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22**(22): 4673-80.

50. **Tjalsma, H. and J. M. van Dijl**. 2005. Proteomics-based consensus prediction of protein retention in a bacterial membrane. Proteomics **5**(17): 4472-82.

51. **Volpe, E., G. Cappelli, M. Grassi, A. Martino, A. Serafino, V. Colizzi, N. Sanarico, and F. Mariani**. 2006. Gene expression profiling of human macrophages at late time of infection with *Mycobacterium tuberculosis*. Immunology **118**(4): 449-60.

52. **Wasinger, V. C. and I. Humphery-Smith**. 1998. Small genes/gene-products in Escherichia coli K-12. FEMS Microbiol Lett **169**(2): 375-82.

53. **Wu, C. H., J. J. Tsai-Wu, Y. T. Huang, C. Y. Lin, G. G. Lioua, and F. J. Lee**. 1998. Identification and subcellular localization of a novel Cu,Zn superoxide dismutase of *Mycobacterium tuberculosis*. FEBS Lett **439**(1-2): 192-6.

54. **Young, D. B. and T. R. Garbe**. 1991. Lipoprotein antigens of *Mycobacterium tuberculosis*. Res Microbiol **142**(1): 55-65.

Chapter VI

Final Discussion and Future Directions

6.1 TB serodiagnosis

Identification of *Mtb* antigens has been achieved through a myriad of experimental techniques, yet few studies have attempted to evaluate antigenic responses on a proteome scale. It was recently shown that different stages of disease were distinguishable via 2-D Western blot analyses of *Mtb* CF proteins (20); however, this methodology is not suitable for high-throughput testing. Advances in protein microarray technology provide a realistic mechanism to screen a large number of serum samples against thousands of proteins to identify biomarkers of disease states. This work established techniques to generate a first generation protein microarray with native CF and cytosolic proteins of *Mtb*. Evaluation of serological reactivity from 42 patients in three tuberculosis disease states and healthy PPD+ individuals demonstrated that HIV negative cavitary- and noncavitary-TB patients recognized 126 and 59 fractions, respectively. Sera from HIV+TB+ patients recognized 20 fractions, of which five overlapped with those recognized by non-HIV TB patients and 15 were unique to the HIV+TB+ disease state. The repertoire of antigens recognized by each disease state confirmed earlier observations that a subset of antigens is recognized early in disease progression (20). The seroreactive components in each fraction were identified with MS-based methods, resulting in 11 previously known antigens and four newly identified antigens. The seroreactivities of recombinant forms of these four antigens were validated by ELISA, bringing this study to a conclusion.

Unique seroreactivity profiles were observed for noncavitary TB patients, cavitary TB patients, and HIV+TB+ patients. These results should enhance our understanding of antigen recognition and disease progression, and this knowledge could assist in developing a future diagnostic assay. The protein microarray specifically described in this work has the capability to assess seroreactivity patterns across other TB disease states. Extrapulmonary TB and pediatric TB contribute significantly to the global TB burden, but have been largely neglected in terms of diagnostic strategies. An understanding of the antigens recognized by these two groups would greatly aid in designing strategies for their diagnosis. Comparative studies are also needed to further address recent proposals that some antigens can differentiate between inactive, latent TB and active TB (4, 5). Testing large numbers of sera banks from different geographic locations should result in further statistical significance and may reveal distinct trends across patient cohorts.

This work revealed a distinct reactivity profile for HIV+TB+ patients. The majority of fractions recognized by this group were not recognized by HIV-negative TB individuals, suggesting aspects of HIV infection result in differential immune recognition of *Mtb* antigens. These disparities in recognition could be due to an altered immune response, atypical disease progression, or modification of bacterial gene expression in response to the host environment. Regardless of the underlying cause, an antigen specific for HIV-coinfection could aid in the development of a targeted diagnostic assay. The similar fractionation characteristics of the 15 unique spots suggested a single antigen may be solely responsible for the unique HIV+TB+ profile. Unfortunately, attempts in identifying this antigen were unsuccessful, and were hindered by limited quantities of

both sample and patients' sera. Future attempts in identifying this antigen will require careful consideration of experimental design. Large amounts of CF proteins will be required as starting material, and fractionation procedures will need to be optimized for isolation of the antigen. Evaluating the reactivity of fractions with HIV+TB+ patients' sera in an ELISA-based format should aid in the purification. Appropriate controls will be essential for this assay, including healthy controls, HIV-TB+, and HIV+TB- patients' sera. Although the identification of an HIV-specific antigen is intriguing, two other antigens, MPT51 and GlcB, were recently shown to provide 91% sensitivity for HIV+TB+ patients (23). In contrast, similar success has not been achieved in identifying antigens capable of diagnosing smear-negative, noncavitary TB patients, suggesting research efforts should instead be more focused on this patient cohort.

The extensive protein fractionation scheme significantly reduced the complexity of cytosolic and CF proteins and permitted the identification of both previously identified and novel antigens. As a case in point, for one particular fraction the novel antigen SodC was enriched for approximately 1000-fold. Furthermore, a relationship was observed where antigens were more often identified in fractions of low complexity than that of high complexity. However, we were unable to identify the reactive components of ~1/3 of the reactive fractions primarily due to insufficient remaining protein quantities. These combined observations suggest the separation scheme walked a fine line between antigen enrichment and sufficient end protein yields for further analysis. As discussed in Chapter 2, alternative fractionation strategies may be expected to result in a different list of identified antigens that displays minimal overlap with the set of antigens identified in this work. However, our work identified 11 previously characterized antigens, with nine of

these repeatedly described in the literature, suggesting our approach was fairly comprehensive. Despite efforts to enrich for proteins of low abundance, we recognize that a relatively small number of novel antigens were identified in this study. An explanation for the low number of newly identified antigens is that the discovery of new serodiagnostic antigens from *in vitro* grown *Mtb* cytosolic and CF protein pools is nearly exhausted. In contrast, the cell envelope protein pool of *M. tuberculosis* is under-exploited for serodiagnostic antigen discovery. Indeed this works suggests the cell envelope may be in fact rich in diagnostic molecules, as each of the four novel antigens described in this work were abundant in the cell wall and membrane fractions, similar to a recent finding regarding *Mtb* T cell antigens (12). The methodologies described in this work lay a foundation for serodiagnostic antigen discovery with the cell envelope fraction, although fractionation of membrane proteins will likely pose unique technical challenges. Future fractionation strategies may consider the use of denaturating agents to further improve protein resolution and increase antigen detection; however, such an approach might also destroy conformational epitopes.

Further *Mtb* antigen discovery will likely be realized once microarray technology is applied to a complete *Mtb* recombinant protein library that potentially represents all 3,989 ORFs. Despite this exciting prospect, it is highly unlikely that a single antigen will be revealed as a "magic bullet", able to diagnose TB disease in all of its manifestation with near 100% accuracy. Thus, a diagnostic assay will probably be based on several antigens with overlapping sensitivities across patient samples. To test this hypothesis, purified forms of the serodiagnostic candidates described thus far could be simultaneously analyzed with the microarray platform. These targeted arrays would have

the potential to identify optimal antigen combinations. Moreover, some antigens (like GlcB) may be useful in identifying all stages of infection, including early infection, while other antigens (like HspX) may be able to fine-tune the prediction and indicate active versus latent infection. These data would make it possible to develop clinical assays based on multiple antigen "barcodes" that will not only enable serodiagnosis of TB, but also the determination of disease severity. Further adding to this approach, a more refined analysis of the top antigens at hand may provide fruitful data. Arrays that display overlapping peptides derived from antigenic proteins may result in the identification of highly reactive epitopes. In addition, peptide arrays should identify non-informative peptides or even epitopes possessing cross-specificity, both of which would contribute to increased background noise in an assay. Development of a clinical assay from the ground up may greatly benefit from inclusion of peptides with high performance that possess defined epitopes and exclusion of peptides that could hinder the assay. Combined with technical improvements such as detection of low titers of antibodies and development of lateral flow and immunochromatographic tests, the identification of optimal antigenic and peptide combinations should provide optimism for a successful serodiagnostic assay in the future.

This work led to the identification of four previously unidentified antigens (BfrB, LppZ, SodC, and TrxC) with potential serodiagnostic utility. Recombinant forms of these antigens were produced in *E. coli* and *M. smegmatis*, and were evaluated in a more traditional ELISA-based format with patients' sera. Each antigen displayed greater than 40% sensitivity in identifying individuals with cavitary TB, thereby validating the methodologies developed in this work that were used for their initial discovery.

However, further testing with larger panels of patients' sera is required to obtain more statistically accurate estimates of their serodiagnostic characteristics, especially regarding their ability to identify individuals with the various clinical manifestations of TB. Notably, noncavitary TB patients' sera were not included in this initial validation study, and serological data from this cohort is needed to compare with the microarray profiles generated from this group.

An interesting finding from this work is that the novel recombinant antigens produced in *M. smegmatis* consistently demonstrated greater seroreactivity than the forms produced in *E. coli*. These differences may suggest *M. smegmatis* is a superior recombinant host for antigen production; however, future studies should be approached with great caution. The differences in seroreactivity are likely due to contaminating *M. smegmatis* seroreactive molecules separate from the recombinant protein of interest. Isolation, identification, production, and characterization of these components across multiple patient cohorts are needed before they are included in a diagnostic assay. Otherwise, similar to crude mycobacterial preparations (3), the presence of undefined components may provide poor specificities.

In the course of validating the seroreactivities of the four novel antigen candidates identified in this work, recombinant forms were produced and analyzed via SDS-PAGE. This led to the observation that the molecular mass of rSodC produced in *M. smegmatis* was larger than that of SP-rSodC produced in *E. coli*. Western blot analysis with the lectin ConA suggested this molecular mass difference was due to glycosylation of rSodC in *M. smegmatis*. However, naturally the question arises as to whether glycosylation of rSodC influences its seroreactivity, as has been shown for MPT32 (13). Indeed, the

ELISA data demonstrated that a greater number of cavitary TB patients' sera recognized rSodC from *M. smegmatis* versus SP-rSodC from *E. coli*; however, each of the recombinant proteins produced in *M. smegmatis* was more seroreactive than their *E. coli* counterpart, and glycosylation alone cannot account for these differences as a whole. Further experimentation suggested SodC glycosylation does not affect its seroreactivity. Western blot analyses with pooled patients' sera indicated no observable differences in the seroreactivities of SP-rSodC from *E. coli,* rSodC from *M. smegmatis*, and rSodC from *Mtb* (data not shown). Furthermore, treatment of rSodC from *Mtb* with α-mannosidase did not coincide with an observable decrease in seroreactivity by Western blot (data not shown). However, these preliminary results require further validation, and analyses with a large panel of individual patient's serum in a more quantitative ELISA format would likely provide more definitive results. Such information could be of importance, as the additional serological recognition of mycobacterial-specific glycosylated epitopes could provide extra sensitivity and/or specificity in a diagnostic assay. The finding that rSodC produced in *M. smegmatis* was glycosylated corroborated a previous observation (9), and provided the opportunity to characterize an *Mtb* glycoprotein with a well-defined function and high-resolution crystal structure. This marked a shift in the research focus of this work, and resulted in a detailed structural analysis of the sites and extent of glycosylation of rSodC produced in *Mtb*.

6.2 Mycobacterial protein glycosylation

Complete structural descriptions have been provided for two mycobacterial glycoproteins, MPT32 and MPB83 (6, 15). This work identified the sites and extent of glycosylation of a third glycoprotein belonging to the MTBC. Similar to MPT32 and

MPB83, SodC possessed multiple O-glycosylated Thr residues localized near the N-terminus that were each modified with a mannose, a mannobiose, or a mannotriose. In addition to Thr O-mannosylation, Ser residues were also shown to be O-mannosylated. The structural characterization of SodC is significant for two reasons. First, this work has effectively doubled the number of identified mycobacterial O-mannosylation sites from six to 12, and this data should aid in the design of glycosylation-site prediction tools specifically trained on mycobacterial O-mannosylation sites. The development of such a tool would greatly aid in defining the repertoire of mycobacterial glycoproteins, an invaluable set of information. If 12 glycosylation sites are not sufficient to build an accurate prediction tool, then the MS-based analyses methods developed in this work should aid in further identifying glycopeptides and O-glycosylated residues within the mycobacterial proteome. Second, the SodC glycoprotein structure provides clues as to the biological significance of mycobacterial protein mannosylation, a concept further discussed below.

An unresolved question from our current studies is the nature of the carbohydrate linkages in SodC. The *Mtb* secreted MPT32 was found to contain α(1→2)-linked mannose residues (6) while the mannobiose of the *M. bovis* lipoprotein MPB83 was found to possess an α(1→3) linkage (15). These linkage disparities may be due to differences in glycoprotein biosynthesis between mycobacterial species, or may be the result of glycosylation differences of lipidated versus nonlipidated secreted proteins. The latter scenario would be especially intriguing, because it would suggest the presence of multiple PMTs with unique specificities, in addition to the established PMT Rv1002c postulated to transfer the initial single mannose residue (25). Linkage analysis of SodC, a

glycosylated *Mtb* lipoprotein, would seem to help address the nature of the linkage differences. However, this work brings into question whether SodC is truly lipidated, and whether the signal peptide is processed with a type I or type II SPase. Therefore, linkage determination of SodC oligomannosides would most likely result in ambiguous interpretations, leaving such questions unanswered. Instead, continued oligomannoside structural analyses of *Mtb* and *M. bovis* confirmed lipoproteins and secreted proteins are needed to address the nature of the differences in carbohydrate linkages.

The identification of glycosylated Ser residues was unexpected, considering that strictly Thr O-glycosylation has been demonstrated in *Mycobacteria* spp. to date (6, 7, 10, 15). This finding raises questions regarding mycobacterial glycoprotein biosynthesis, namely whether a single mycobacterial PMT can recognize and glycosylate Thr and Ser residues, or whether separate PMTs are required for each hydroxy amino acid. The future identification and analyses of additional PMTs from *Mtb* should aid in the understanding of PMT substrate specificity. However, the specificity of the PMT Rv1002c could readily be determined with an experimental procedure similar to that described by VanderVen *et al.* (25). Specifically, synthetic peptide acceptors could be designed that are based on the SodC N-terminal glycosylated sequence with and without Thr residues substituted with Ala residues. Mycobacterial cytoplasmic membrane extracts expressing recombinant *rv1002c* could then be assayed for the ability to transfer radiolabelled mannose from GDP-[^{14}C]-mannose onto the various peptide acceptors, and in this manner the ability of Rv1002c to transfer mannose to Ser residues could be determined. Comparing the relative PMT activities of membrane extracts from a strain

expressing *rv1002c* versus a strain containing an empty vector would be necessary to establish Rv1002c substrate specificity.

The function of protein glycosylation in mycobacteria is an enigma. Protein mannosylation has been shown to influence immune recognition (11, 13, 19) and interactions with the host cell (18). Although such effects on *Mtb* pathogenesis are intriguing, protein glycosylation likely serves a more fundamental role in the physiology of the bacillus. This hypothesis is supported by the essentiality of the PMT Rv1002c gene product for *in vitro* growth as proposed by transposon mutagenesis studies (22). Furthermore numerous studies, including this work, have shown *Mtb* protein sequences to be glycosylated when produced in *M. smegmatis* (2, 8, 9), demonstrating the presence of intact protein mannosylation systems in both pathogenic and nonpathogenic mycobacteria. In addition to providing the mannoprotein structure of SodC, this work resulted in two important findings that should contribute to the future elucidation of the primary role of mycobacterial protein glycosylation.

The first finding suggests glycosylation likely does not regulate the catalytic activity of SodC. This hypothesis is based on a combination of SodC protein sequence comparisons across diverse species and the likely physical separation of the unfolded, extended glycosylated sequence from the previously determined folded enzymatic SOD domain (24). Future SOD quantitative activity-based experiments are needed to validate this hypothesis and to explore the potential effects glycosylation has on cofactor usage or protein dimerization, and the various mutated and non-mutated rSodC expression constructs described in this work should greatly aid in such experiments. Furthermore, as crystal structures and functional domains of *Mtb* proteins continue to be defined, future

comparisons with mycobacterial glycoprotein sequences, identified experimentally or through advanced glycoprotein prediction tools, should help in addressing the relationship between glycosylation and enzymatic activity.

The second finding adds to growing evidence that suggests mycobacterial protein glycosylation may be directing proteolytic cleavage of lipoproteins. Several observations support this hypothesis. We identified processed forms of SodC in the CF, and this was similar to previous findings for two other glycolipoproteins, in which truncated forms of the 38-kDa antigen and MPB83 were also observed in this fraction (1, 15). SodC mannosylation is localized near the N-terminus, in general agreement with the predicted location of glycosylation for most *Mtb* lipoproteins as determined through bioinformatic analyses (9). Similar to a report for the 19-kDa lipoprotein (10), substitution of a SodC glycosylated Thr cluster in the N-terminal region resulted in truncation of the N-terminus. Thus, SodC data supports any of the three previous previously proposed mechanisms, 1) glycosylation serves as a signal for cleavage, 2) glycosylation protects from further amino-terminal degradation following proteolytic cleavage from the acylated anchor, or 3) glycosylation protects lipoproteins from initial proteolytic cleavage of the acylated anchor (10, 15). Recently, parallels were drawn between mycobacterial and yeast mannoprotein biosynthesis (25). Therefore, it may be hypothesized that mycobacterial mannosylation serves similar functions to yeast O-glycosylation, which include protection from proteolytic processing (14, 21) and proper protein localization (17).

The *in vivo* contribution of SodC glycosylation could be analyzed in two different manners. Firstly, Piddington *et al.* (16) showed that an *Mtb sodC* null mutant was readily killed in broth cultures supplemented with the extracellular superoxide-generating

compounds hypoxanthine and xanthine oxidase. Thus, complementation of this mutant with the non-mutated *sodC* and the various *sodC* mutants encoding Thr to Ala substitutions could result in different survival phenotypes, suggesting glycosylation affects the function of SodC. Secondly, perhaps of even greater interest is the finding that an *Mtb sodC* null mutant showed increased susceptibility to killing by IFN-γ-activated murine macrophages (16). Complementation with chromosomal copies of non-mutated *sodC* and the various *sodC* mutants encoding Thr to Ala substitutions could demonstrate that *Mtb* protein glycosylation contributes to virulence, a significant finding. Although both of these experiments could establish *in vivo* roles for protein glycosylation in mycobacterial defense, these methods would not resolve the manner in which glycosylation affects SodC function.

The proteolytic processing hypothesis supports an overall model where mycobacterial protein glycosylation serves to regulate the turnover of cell wall proteins and lipoproteins. By varying the extent of glycosylation, *Mtb* may be able to finely modulate the placement and activity of enzymes during different growth phases of the bacillus. Recent insight gained into the biosynthesis of *Mtb* mannoproteins (25) along with the tools and information from this current work will allow careful testing of this hypothesis. Identification of the key players in these processes and elucidation of their interactions could provide novel drug targets and a better understanding of mycobacterial physiology and pathogenesis in general.

6.3 Literature cited

1. **Andersen, A. B., L. Ljungqvist, and M. Olsen**. 1990. Evidence that protein antigen b of *Mycobacterium tuberculosis* is involved in phosphate metabolism. J Gen Microbiol **136**(3): 477-80.

2. **Cooper, H. N., S. S. Gurcha, J. Nigou, P. J. Brennan, J. T. Belisle, G. S. Besra, and D. Young**. 2002. Characterization of mycobacterial protein glycosyltransferase activity using synthetic peptide acceptors in a cell-free assay. Glycobiology **12**(7): 427-34.

3. **Daniel, T. M. and S. M. Debanne**. 1987. The serodiagnosis of tuberculosis and other mycobacterial diseases by enzyme-linked immunosorbent assay. Am Rev Respir Dis **135**(5): 1137-51.

4. **Davidow, A., G. V. Kanaujia, L. Shi, J. Kaviar, X. Guo, N. Sung, G. Kaplan, D. Menzies, and M. L. Gennaro**. 2005. Antibody profiles characteristic of *Mycobacterium tuberculosis* infection state. Infect Immun **73**(10): 6846-51.

5. **Demissie, A., E. M. Leyten, M. Abebe, L. Wassie, A. Aseffa, G. Abate, H. Fletcher, P. Owiafe, P. C. Hill, R. Brookes, G. Rook, A. Zumla, S. M. Arend, M. Klein, T. H. Ottenhoff, P. Andersen, and T. M. Doherty**. 2006. Recognition of stage-specific mycobacterial antigens differentiates between acute and latent infections with *Mycobacterium tuberculosis*. Clin Vaccine Immunol **13**(2): 179-86.

6. **Dobos, K. M., K. H. Khoo, K. M. Swiderek, P. J. Brennan, and J. T. Belisle**. 1996. Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. J Bacteriol **178**(9): 2498-506.

7. **Dobos, K. M., K. Swiderek, K. H. Khoo, P. J. Brennan, and J. T. Belisle**. 1995. Evidence for glycosylation sites on the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. Infect Immun **63**(8): 2846-53.

8. **Garbe, T., D. Harris, M. Vordermeier, R. Lathigra, J. Ivanyi, and D. Young**. 1993. Expression of the *Mycobacterium tuberculosis* 19-kilodalton antigen in *Mycobacterium smegmatis*: immunological analysis and evidence of glycosylation. Infect Immun **61**(1): 260-7.

9. **Herrmann, J. L., R. Delahay, A. Gallagher, B. Robertson, and D. Young**. 2000. Analysis of post-translational modification of mycobacterial proteins using a cassette expression system. FEBS Lett **473**(3): 358-62.

10. **Herrmann, J. L., P. O'Gaora, A. Gallagher, J. E. Thole, and D. B. Young**. 1996. Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from *Mycobacterium tuberculosis*. Embo J **15**(14): 3547-54.

11. **Horn, C., A. Namane, P. Pescher, M. Riviere, F. Romain, G. Puzo, O. Barzu, and G. Marchal**. 1999. Decreased capacity of recombinant 45/47-kDa molecules (Apa) of *Mycobacterium tuberculosis* to stimulate T lymphocyte responses related to changes in their mannosylation pattern. J Biol Chem **274**(45): 32023-30.

12. **Laal, S. and Y. A. Skeiky**, Immune-Based Methods, in Tuberculosis and the Tubercle Bacillus, S.T. Cole, Editor. 2005, ASM Press: Washington, D.C. p. 71-83.

13. **Lara, M., L. Servin-Gonzalez, M. Singh, C. Moreno, I. Cohen, M. Nimtz, and C. Espitia**. 2004. Expression, secretion, and glycosylation of the 45- and 47-kDa glycoprotein of *Mycobacterium tuberculosis* in *Streptomyces lividans*. Appl Environ Microbiol **70**(2): 679-85.

14. **Lommel, M., M. Bagnat, and S. Strahl**. 2004. Aberrant processing of the WSC family and Mid2p cell surface sensors results in cell death of Saccharomyces cerevisiae O-mannosylation mutants. Mol Cell Biol **24**(1): 46-57.

15. **Michell, S. L., A. O. Whelan, P. R. Wheeler, M. Panico, R. L. Easton, A. T. Etienne, S. M. Haslam, A. Dell, H. R. Morris, A. J. Reason, J. L. Herrmann, D. B. Young, and R. G. Hewinson**. 2003. The MPB83 antigen from *Mycobacterium bovis* contains O-linked mannose and (1-->3)-mannobiose moieties. J Biol Chem **278**(18): 16423-32.

16. **Piddington, D. L., F. C. Fang, T. Laessig, A. M. Cooper, I. M. Orme, and N. A. Buchmeier**. 2001. Cu,Zn superoxide dismutase of *Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst. Infect Immun **69**(8): 4980-7.

17. **Proszynski, T. J., K. Simons, and M. Bagnat**. 2004. O-glycosylation as a sorting determinant for cell surface delivery in yeast. Mol Biol Cell **15**(4): 1533-43.

18. **Ragas, A., L. Roussel, G. Puzo, and M. Riviere**. 2007. The *Mycobacterium tuberculosis* cell-surface glycoprotein apa as a potential adhesin to colonize target cells via the innate immune system pulmonary C-type lectin surfactant protein A. J Biol Chem **282**(8): 5133-42.

19. **Romain, F., C. Horn, P. Pescher, A. Namane, M. Riviere, G. Puzo, O. Barzu, and G. Marchal**. 1999. Deglycosylation of the 45/47-kilodalton antigen complex of *Mycobacterium tuberculosis* decreases its capacity to elicit in vivo or in vitro cellular immune responses. Infect Immun **67**(11): 5567-72.

20. **Samanich, K., J. T. Belisle, and S. Laal**. 2001. Homogeneity of antibody responses in tuberculosis patients. Infect Immun **69**(7): 4600-9.

21. **Sanders, S. L., M. Gentzsch, W. Tanner, and I. Herskowitz**. 1999. O-Glycosylation of Axl2/Bud10p by Pmt4p is required for its stability, localization, and function in daughter cells. J Cell Biol **145**(6): 1177-88.

22. **Sassetti, C. M., D. H. Boyd, and E. J. Rubin**. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol **48**(1): 77-84.

23. **Singh, K. K., Y. Dong, J. T. Belisle, J. Harder, V. K. Arora, and S. Laal.** 2005. Antigens of *Mycobacterium tuberculosis* recognized by antibodies during incipient, subclinical tuberculosis. Clin Diagn Lab Immunol **12**(2): 354-8.

24. **Spagnolo, L., I. Toro, M. D'Orazio, P. O'Neill, J. Z. Pedersen, O. Carugo, G. Rotilio, A. Battistoni, and K. Djinovic-Carugo.** 2004. Unique features of the sodC-encoded superoxide dismutase from *Mycobacterium tuberculosis*, a fully functional copper-containing enzyme lacking zinc in the active site. J Biol Chem **279**(32): 33447-55.

25. **VanderVen, B. C., J. D. Harder, D. C. Crick, and J. T. Belisle.** 2005. Export-mediated assembly of mycobacterial glycoproteins parallels eukaryotic pathways. Science **309**(5736): 941-3.