Recent advances in defining the immunoproteome of *Mycobacterium tuberculosis*

Shajo Kunnath-Velayudhan[1] and Steven Anthony Porcelli[1,2]*

[1] Department of Microbiology and Immunology, Albert Einstein College of Medicine of Yeshiva University, Bronx, NY, USA
[2] Department of Medicine, Albert Einstein College of Medicine of Yeshiva University, Bronx, NY, USA

Immunity conferred by antigen-specific CD4+ T cells is critical for controlling infection with *Mycobacterium tuberculosis* (Mtb), the causative agent of tuberculosis. However, despite research that spans more than a century, many of the characteristics of protective immune responses to Mtb remain elusive. Defining the repertoire of antigenic targets is central to understanding the immune response against this pathogen. Although traditional methods of antigen discovery have identified many immunodominant antigens, they afford limited proteome coverage. Recent advances in proteomic techniques that are based on peptide library and protein microarray technology have enabled interrogation of the entire proteome of Mtb for antigens. Though these techniques have limitations and are still evolving, early studies using these techniques provide an unbiased view of the immune response to Mtb. Here we review proteome-wide approaches to antigen discovery and summarize what these have revealed so far on the composition of the Mtb immunoproteome.

Keywords: peptide library, protein microarray, antigen discovery, type VII secretion system, ESX proteins, PE/PPE proteins
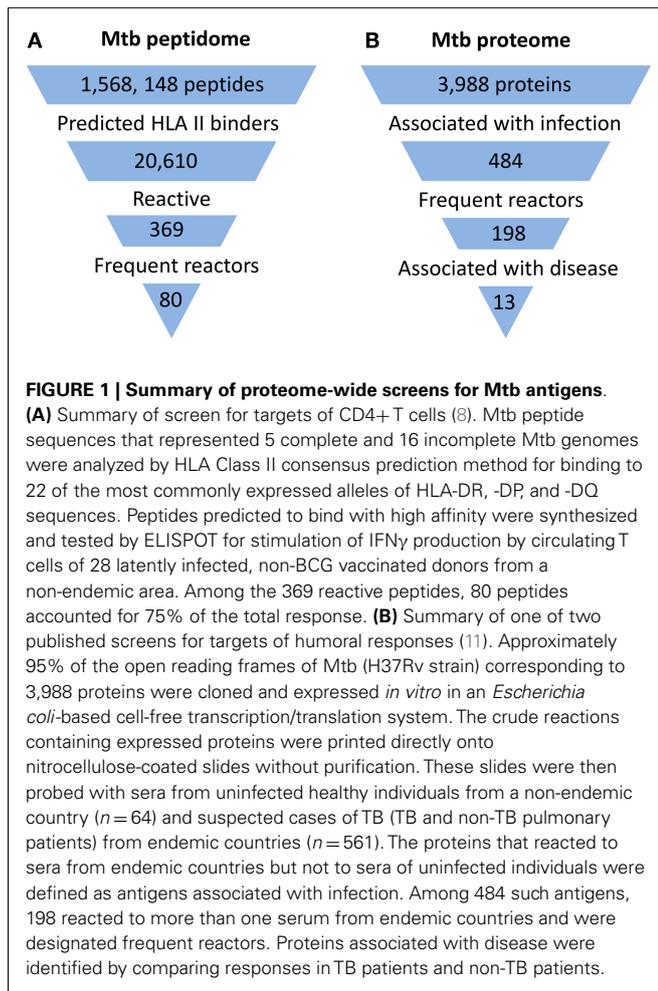
INTRODUCTION

Decades of research on T cell responses to *Mycobacterium tuberculosis* (Mtb) have identified many immunodominant antigens, some of which provide significant protection as immunogens in animal models (1). A major focus has been to identify antigens recognized by CD4+ T cells, since these cells are believed to play a dominant role in controlling Mtb infection (2). However, robust immune responses induced by immunodominant antigens do not necessarily translate into protective immunity (3), which reflects the complex and incompletely understood role of host immunity in the natural history of tuberculosis. On the one hand, antigen-specific CD4+ T cell responses are critical for protection against tuberculosis, and Mtb has evolved many strategies that subvert and evade the host adaptive immune response (4). On the other hand, Mtb exploits immune responses for its own benefit, and evidence suggests that immune-mediated tissue destruction facilitates the spread of Mtb among hosts. In agreement with this, recent studies have found that T cell epitopes of known immunodominant antigens of Mtb are hyperconserved, implying that immune responses against them may be in some cases more beneficial to the bacilli than to the host (5).

Given this complex relationship between host and pathogen, a full definition of the antigenic repertoire or "immunoproteome" of Mtb is an important step toward understanding how to effectively vaccinate against this infection. Until recently, most approaches to antigen discovery were based on traditional methods for separation and identification of antigens from complex mycobacterial protein mixtures (6). Alternative methods have employed screening of Mtb expression libraries in *Escherichia coli* with T cell clones derived from latently infected individuals (7). Although these methods have been successful in identifying immunodominant

antigens, they provide restricted coverage of the Mtb proteome which has over 4,000 proteins. Advances in proteome-wide screening methods now enable a more extensive and unbiased survey of antigenic targets on complex pathogens like Mtb. Here, we review results from recently published proteome-wide antigen screens, and discuss how this powerful new approach may improve our understanding of the CD4+ T cell response to Mtb.

DEFINING THE IMMUNOPROTEOME OF Mtb

With the development of technologies that allow high-throughput peptide and protein synthesis, it is now possible to interrogate the entire Mtb proteome for antigens. So far, three proteome-wide and relatively unbiased approaches to identify candidate antigens for CD4+ T cell responses from the Mtb proteome have been described (**Figure 1**). One approach was based on the use of a peptide library designed to screen potential targets of CD4+ T cell responses in latently infected individuals (8). In contrast to previous studies which typically involved *in vitro* expansion of Mtb-specific T cells, circulating T cells from Mtb infected donors were directly tested against the synthetic peptide library using IFNγ ELISPOT assay. This screen involved predicting Mtb peptides that bind with high affinity to commonly expressed MHC class II alleles using a consensus approach based on results from three prediction methods (9). Two other studies used analysis of serum antibody responses as a surrogate for CD4+ T cell responses, relying on the assumption that a strong linkage exists between the targets of antibodies and of the CD4+ helper T cells involved in their generation (10). In one case, protein microarrays printed with products of all expressed open reading frames of Mtb were used to screen sera from TB patients and controls for antibody reactivity (11). In the other study, a similar screen for

FIGURE 1 | Summary of proteome-wide screens for Mtb antigens.
(A) Summary of screen for targets of CD4+ T cells (8). Mtb peptide
sequences that represented 5 complete and 16 incomplete Mtb genomes
were analyzed by HLA Class II consensus prediction method for binding to
22 of the most commonly expressed alleles of HLA-DR, -DP, and -DQ
sequences. Peptides predicted to bind with high affinity were synthesized
and tested by ELISPOT for stimulation of IFNγ production by circulating T
cells of 28 latently infected, non-BCG vaccinated donors from a
non-endemic area. Among the 369 reactive peptides, 80 peptides
accounted for 75% of the total response. **(B)** Summary of one of two
published screens for targets of humoral responses (11). Approximately
95% of the open reading frames of Mtb (H37Rv strain) corresponding to
3,988 proteins were cloned and expressed *in vitro* in an *Escherichia
coli*-based cell-free transcription/translation system. The crude reactions
containing expressed proteins were printed directly onto
nitrocellulose-coated slides without purification. These slides were then
probed with sera from uninfected healthy individuals from a non-endemic
country (*n* = 64) and suspected cases of TB (TB and non-TB pulmonary
patients) from endemic countries (*n* = 561). The proteins that reacted to
sera from endemic countries but not to sera of uninfected individuals were
defined as antigens associated with infection. Among 484 such antigens,
198 reacted to more than one serum from endemic countries and were
designated frequent reactors. Proteins associated with disease were
identified by comparing responses in TB patients and non-TB patients.

serum antibody responses against the Mtb proteome in TB patients
was performed using traditional methods of recombinant protein
expression and ELISA (12).

Despite major differences in their goals and methodology, all
studies concluded that human immune responses target a sub-
set of the Mtb proteome during infection (**Figure 1**). In the direct
screen for CD4+ T cell responses to synthetic peptides, ~2% of the
proteome accounted for 80% of the total responses (8). With the
serum antibody-based screens, a larger fraction of the proteome
was found to be immunogenic, constituting ~6–10% of the pro-
teome (11, 12). In part, this difference may reflect that the latter two
studies screened for immune responses in active TB patients, who
would generally be expected to have higher bacterial burdens than
the latently infected individuals studied by Lindestam Arlehamn
et al. A striking feature of all studies was that the subset of the
proteome that induced an immune response (i.e., the immunopro-
teome) was enriched for secreted and cell wall-associated proteins.
This is consistent with earlier studies which have suggested that the
immunodominant antigens of mycobacteria are proteins that are
secreted by the bacteria during infection, and also with the recent
finding that mycobacteria secrete membrane vesicles rich in anti-
gens (13). Although the components of the immunoproteome

varied and showed only partial overlap among these studies, a
consistent finding was the remarkable prominence of immune
responses directed at members of the PE/PPE and ESX protein
families, which are well known as critical virulence determinants
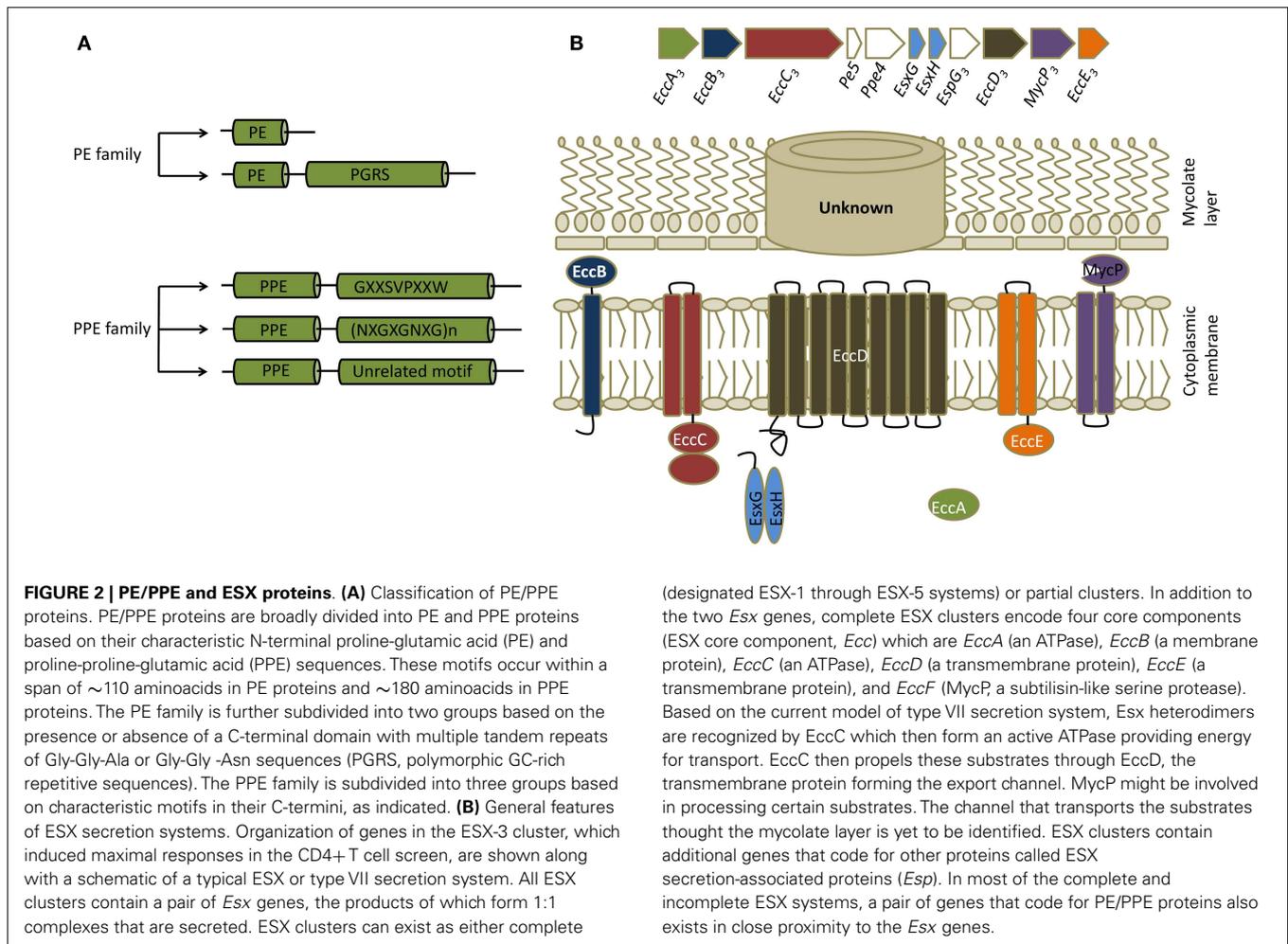of mycobacteria.

PE/PPE PROTEINS AS TARGETS OF CD4+ T CELLS
PE/PPE proteins are unique proteins found exclusively in path-
ogenic mycobacteria. There are genes for 99 PE proteins and
68 PPE proteins in the genome of Mtb, constituting ~10% of
its coding capacity (14). They are characterized by the presence
of proline-glutamic acid (PE) and proline-proline-glutamic acid
(PPE) motifs near their N-termini (15). These proteins are further
divided into subfamilies based on the motifs in their C-termini
(**Figure 2**). Many PE/PPE proteins are localized to the bacterial
cell surface and many are secreted, making them readily accessible
to the immune system (16). Indeed, PE/PPE proteins are known
to induce strong humoral and cellular immune responses based
on studies that used traditional immunological assays (17). In the
genome, PE/PPE genes are often found within operons or gene
clusters that also contain a variety of other types of genes. Most
notably, genes for some PE/PPE proteins map within gene clusters
encoding type VII secretion systems, which are known as ESX sys-
tems in mycobacteria (see below). Secretion of some of the PE/PPE
proteins is known to be dependent on ESX systems, and compar-
ative genomics studies suggest that ESX clusters have coevolved
with PE/PPE proteins. The role of this distinct family of proteins
in mycobacterial virulence is yet to be determined, although stud-
ies suggest involvement in many aspects of pathogenesis including
bacterial attachment to host cells, immunomodulation, and ability
to persist in granulomas (15, 17).

All three published proteome-wide antigen screens showed that
PE/PPE proteins are major targets of human immune responses.
Among the antigens identified by peptide library screening for
CD4+ T cell responses, ~45% were PE/PPE proteins though
they constitute <5% of the open reading frames of Mtb. In
addition, PE/PPE proteins represented approximately half of the
novel antigens identified, suggesting that immune responses to
these proteins remain mostly unexplored. Proteome-wide anti-
body screening also revealed an enrichment of PE/PPE proteins
among the antigens though the enrichment was less than that
observed for CD4 responses (an average of ~5–6% of the total
targets recognized). It is possible that the relatively lower enrich-
ment found for these proteins for antibody responses compared
to cellular responses could in part reflect difficulties in expressing
these proteins by recombinant techniques that were used to gener-
ate antigens for the antibody-based screening. The identification
of PE/PPE proteins as prominent antigenic targets of CD4+ T
cells in these studies of the immunoproteome is consistent with
other more targeted analyses showing that these proteins are highly
immunogenic in mycobacteria-infected humans and cattle (18).

COMPONENTS OF ESX SECRETION SYSTEM AS TARGETS OF CD4+ T CELLS
ESX secretion systems are specialized protein export systems orig-
inally identified in mycobacteria and constitute a distinct mecha-
nism for protein secretion known as the type VII secretion system.

FIGURE 2 | PE/PPE and ESX proteins. **(A)** Classification of PE/PPE proteins. PE/PPE proteins are broadly divided into PE and PPE proteins based on their characteristic N-terminal proline-glutamic acid (PE) and proline-proline-glutamic acid (PPE) sequences. These motifs occur within a span of ~110 aminoacids in PE proteins and ~180 aminoacids in PPE proteins. The PE family is further subdivided into two groups based on the presence or absence of a C-terminal domain with multiple tandem repeats of Gly-Gly-Ala or Gly-Gly -Asn sequences (PGRS, polymorphic GC-rich repetitive sequences). The PPE family is subdivided into three groups based on characteristic motifs in their C-termini, as indicated. **(B)** General features of ESX secretion systems. Organization of genes in the ESX-3 cluster, which induced maximal responses in the CD4+ T cell screen, are shown along with a schematic of a typical ESX or type VII secretion system. All ESX clusters contain a pair of *Esx* genes, the products of which form 1:1 complexes that are secreted. ESX clusters can exist as either complete

(designated ESX-1 through ESX-5 systems) or partial clusters. In addition to the two *Esx* genes, complete ESX clusters encode four core components (ESX core component, *Ecc*) which are *EccA* (an ATPase), *EccB* (a membrane protein), *EccC* (an ATPase), *EccD* (a transmembrane protein), *EccE* (a transmembrane protein), and *EccF* (MycP, a subtilisin-like serine protease). Based on the current model of type VII secretion system, Esx heterodimers are recognized by EccC which then form an active ATPase providing energy for transport. EccC then propels these substrates through EccD, the transmembrane protein forming the export channel. MycP might be involved in processing certain substrates. The channel that transports the substrates thought the mycolate layer is yet to be identified. ESX clusters contain additional genes that code for other proteins called ESX secretion-associated proteins (*Esp*). In most of the complete and incomplete ESX systems, a pair of genes that code for PE/PPE proteins also exists in close proximity to the *Esx* genes.

The system is named after two proteins, EsxA (also known as Early Secretory Antigenic Target-6, or ESAT-6) and EsxB (Culture Filtrate Protein-10, or CFP-10). There are five complete (ESX-1–5) and five to six incomplete ESX secretion systems in Mtb, and all carry a pair of genes that encode homologs of EsxA and EsxB (**Figure 2**). Most of the complete and incomplete ESX systems also contain or are closely linked to a pair of genes that code for PE/PPE proteins. These secretion systems export small proteins that contain a WXG amino acid motif (tryptophan and glycine separated by one amino acid). Secretion of EsxA and EsxB or their homologs, as well as many PE/PPE proteins, is dependent on ESX systems. Despite similarities, ESX systems do not complement each other and each one is likely to play distinct roles in Mtb virulence and physiology. For example, ESX-1 is required for survival of mycobacteria in mice, for granuloma formation, and for escape of Mtb from phagosome into cytosol (19, 20). Similarly, ESX-3 is involved in iron and zinc uptake by the bacteria, while ESX-5 has a role in modulating immune responses (20). It should be noted that these systems and their homologs are found in pathogenic and non-pathogenic mycobacteria and in other gram positive bacteria (21).

ESX proteins are among the well known immunodominant antigens of Mtb. Commercial tests for Mtb infection (QuantiF-ERON® and T-SPOT.TB) use EsxA and EsxB as antigens since circulating T cells from most infected individuals respond strongly to these proteins. However, the proteome-wide screens reveal for the first time the extent of immunodominance of ESX proteins. For example, Lindestam Arlehamn et al. showed that 42% of the cellular responses they detected were directed against a set of proteins that represented 0.55% of the Mtb genome. These proteins were encoded by genes located in three distinct regions of the Mtb genome, which the authors called antigenic islands. All three islands include Esx protein pairs (i.e., EsxA/EsxB or homologs) and two contain the complete type VII secretion systems ESX-1 and ESX-3. Immunoproteome analysis using antibody screening also revealed that components of ESX systems induce strong immune responses in humans. Pathway analysis of the data from probing of microarrays with sera from active TB patients showed a significant enrichment in reactivity to proteins controlled by Zur, a regulator of Zinc uptake. Zur regulates a set of 56 genes which includes genes that code for three ESX systems. Interestingly, two of these three systems (ESX-3 and a partial ESX cluster containing

genes for EsxQ, EsxR, and EsxS) were included in the three anti-genic islands identified in the proteome-wide screen for CD4+ T cell responses. In addition, two proteins of the ESX-1 system were among the top five antigens associated with active TB. Collectively, these data underscore the immunodominance of ESX systems.

Surprisingly, it appears that both secreted and non-secreted components of the ESX systems induce strong immune responses. For example, previous studies have shown that eight different T cell antigens are encoded by genes in and around the ESX-1 region. Since the antigens tend to cluster in the genome, ESX systems had been called immunogenicity islands (22). It is not clear how the non-secreted components of the ESX secretion systems induce strong immune responses compared to other unrelated, surface-associated proteins. Since homologs of ESX systems exist in many bacteria, it is possible that cross-reactive epitopes may contribute to enhanced recognition of non-secreted components. Co-regulation might also explain some of the immunodominance. For example, bacteria may overexpress secreted and non-secreted components of the ESX-3 system in conditions of low iron or zinc. In addition, the components of these systems might exist as multi-protein aggregates, as shown for the ESX-5 system (23). This could favor the spreading of immune responses from epitopes of the secreted components to epitopes of the non-secreted components. Another possibility is that the secretion system may form structures protruding from the bacterial cell surface, thus increasing accessibility to antigen processing machinery (19). Although the immunoproteome analysis suggested that ESX proteins, and less strikingly PE/PPE proteins, in the antigenic islands were more immunogenic than their counterparts encoded in other areas of the genome (8), other more focused studies in *M. bovis* infected cattle have not confirmed this finding (18, 22, 24).

LIMITATIONS OF CURRENT METHODS FOR PROTEOME-WIDE SCREENING

Although uniquely powerful, high-throughput approaches of antigen discovery do not necessarily capture responses to all immunodominant antigens. For example, in the screen for cellular responses, only 20–25% of the latently infected donors responded to EsxA or EsxB (8). These antigens are known to induce responses in most latently infected individuals, which is the basis for their use in commercial *ex vivo* tests for latent infection. Similarly, in another analysis of 18 latently infected individuals based on responses to peptide libraries (25), no responses were detected to the TB7.7 antigen. This antigen is also used in a commercial test for latent infection and stimulates responses in approximately half of infected individuals (26, 27). Similarly, proteome-wide screens based on expression of recombinant proteins may yield false negative results because of failure to express certain Mtb proteins. Indeed, the microarray-based screen for humoral responses failed to detect some known immunodominant antigens, importantly EsxA (11). In addition, the screen by Li et al. failed to detect responses to both EsxA and EsxB. Another limitation is that pre-selection of peptides that are predicted to bind to MHC class II molecules by Lindestam Arlehamn et al. may have excluded some immunodominant peptides, since the performance of the predictive algorithms used is known to suffer from a significant rate of both false positive and false negative errors (9).

CONSERVATION VERSUS VARIATION OF EPITOPE SEQUENCES

The question of whether certain immune responses induced by Mtb antigens could actually be more beneficial to the bacteria than to the host remains open. According to classical models that portray the host-pathogen interaction as an "evolutionary arms race," immunodominant epitopes are likely to be less conserved due to high selection pressure from the immune response. Recent studies suggest that this model may not consistently apply to mycobacteria. By analyzing genome sequences of a diverse collection of Mtb strains, Comas et al. showed that the great majority of known T cell epitopes display little sequence variation, with lower ratios of non-synonymous to synonymous changes than seen in other coding regions of essential and non-essential genes (5). Though the study by Comas et al. excluded members of PE/PPE protein family for technical reasons, a recent study of Mtb isolates showed that PE/PPE proteins are also not under diversifying selection pressure (28). One caveat of these studies is that they used epitope sequences that were originally identified from studies that used the genomic sequence of a single laboratory strain of Mtb (H37Rv), and this approach would be expected not to include epitopes that are highly variable among random clinical isolates. Thus, the analysis may have been inadvertently restricted to invariant epitopes, leading to severe biasing of the results. In addition, use of a small number of antigens might have caused further bias, as approximately two thirds of the epitopes in the database used to design these studies (the Immune Epitope Database) are derived from only thirty antigens (29). In contrast to the studies showing conservation of epitope sequences, another study has shown evidence for sequence variation in known T cell epitopes suggesting that they may diversify in order to evade immune recognition (30). Thus, a clear and consistent picture of the level of epitope conservation versus variation has yet to be established.

For antigens that derive from multigene families such as the PE/PPE and ESX proteins, another potentially important feature that may contribute to immunodominance is the presence of cross-reactive epitopes that are shared by multiple homologous family members. Indeed, PE/PPE proteins are known to carry highly conserved sequences. The relevance of this point was suggested by a study showing that both magnitude and frequency of T cell responses to PE/PPE proteins were greater to peptide pools representing the more conserved N-terminal regions than to pools representing other regions of the proteins (18). In addition, a majority of the reactive peptides showed more than 70% sequence similarity to one or more additional PE/PPE regions, further suggesting that cross reactivity contributes significantly to responses to PE/PPE proteins. Another study which examined responses to ESX-5 encoded PE/PPE proteins in mice also concluded that cross-reactive epitopes contribute significantly to the immunodominance of PE/PPE proteins (31). However, in contrast to the response to PE/PPE proteins, analysis of bovine responses to ESX proteins did not reveal an association between immunodominance of specific epitopes and their level of cross reactivity with other homologous Esx proteins (24).

CONCLUSION

New methods enabling proteome-wide antigen screens provide for the first time a general view of the immune responses induced by Mtb. The data obtained so far using these methods indicate that during Mtb infection, immune responses target a small subset of the proteome enriched for membrane-associated and secreted proteins. Among these antigens, members of the families of PE/PPE and ESX proteins are major targets of immune responses. These results should encourage future efforts to characterize the role of these protein families in pathogenesis, and stimulate interest in developing them as components of novel vaccines. While these two protein families comprise important potential components of novel vaccines, it remains to be determined whether the most prominent antigens in natural infection are also among the most effective antigens in the context of vaccination. Since most of the data obtained so far have come from analyses of immune responses in chronically infected individuals, it will be important to extend these studies to subjects with acute infection who may target a different antigenic repertoire.

REFERENCES

1. McShane H. Tuberculosis vaccines: beyond bacille Calmette-Guerin. *Philos Trans R Soc Lond B Biol Sci* (2011) **366**:2782–9. doi:10.1098/rstb.2011.0097

2. Flynn JL, Chan J. Immunology of tuberculosis. *Annu Rev Immunol* (2001) **19**:93–129. doi:10.1146/annurev.immunol.19.1.93

3. Tameris MD, Hatherill M, Landry BS, Scriba TJ, Snowden MA, Lockhart S, et al. Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. *Lancet* (2013) **381**:1021–8. doi:10.1016/S0140-6736(13)60177-4

4. Baena A, Porcelli SA. Evasion and subversion of antigen presentation by *Mycobacterium tuberculosis*. *Tissue Antigens* (2009) **74**:189–204. doi:10.1111/j.1399-0039.2009.01301.x

5. Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet* (2010) **42**:498–503. doi:10.1038/ng.590

6. Andersen P, Doherty TM. TB subunit vaccines – putting the pieces together. *Microbes Infect* (2005) **7**:911–21. doi:10.1016/j.micinf.2005.03.013

7. Coler RN, Dillon DC, Skeiky YA, Kahn M, Orme IM, Lobet Y, et al. Identification of *Mycobacterium tuberculosis* vaccine candidates using human CD4+ T-cells expression cloning. *Vaccine* (2009) **27**:223–33. doi:10.1016/j.vaccine.2008.10.056

8. Lindestam Arlehamn CS, Gerasimova A, Mele F, Henderson R, Swann J, Greenbaum JA, et al. Memory T cells in latent *Mycobacterium tuberculosis* infection are directed against three antigenic islands and largely contained in a CXCR3+CCR6+ Th1 subset. *PLoS Pathog* (2013) **9**:e1003130. doi:10.1371/journal.ppat.1003130

9. Wang P, Sidney J, Dow C, Mothe B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol* (2008) **4**:e1000048. doi:10.1371/journal.pcbi.1000048

10. Sette A, Moutaftsi M, Moyron-Quiroz J, McCausland MM, Davies DH, Johnston RJ, et al. Selective CD4+ T cell help for antibody responses to a large viral pathogen: deterministic linkage of specificities. *Immunity* (2008) **28**:847–58. doi:10.1016/j.immuni.2008.04.018

11. Kunnath-Velayudhan S, Salamon H, Wang HY, Davidow AL, Molina DM, Huynh VT, et al. Dynamic antibody responses to the *Mycobacterium tuberculosis* proteome. *Proc Natl Acad Sci USA* (2010) **107**:14703–8. doi:10.1073/pnas.1009080107

12. Li Y, Zeng J, Shi J, Wang M, Rao M, Xue C, et al. A proteome-scale identification of novel antigenic proteins in *Mycobacterium tuberculosis* toward diagnostic and vaccine development. *J Proteome Res* (2010) **9**:4812–22. doi:10.1021/pr1005108

13. Prados-Rosales R, Baena A, Martinez LR, Luque-Garcia J, Kalscheuer R, Veeraraghavan U, et al. Mycobacteria release active membrane vesicles that modulate immune responses in a TLR2-dependent manner in mice. *J Clin Invest* (2011) **121**:1471–83. doi:10.1172/JCI44261

14. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* (1998) **393**:537–44. doi:10.1038/31159

15. Sampson SL. Mycobacterial PE/PPE proteins at the host-pathogen interface. *Clin Dev Immunol* (2011) **2011**:497203. doi:10.1155/2011/497203

16. Abdallah AM, Verboom T, Weerdenburg EM, Gey van Pittius NC, Mahasha PW, Jimenez C, et al. PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. *Mol Microbiol* (2009) **73**:329–40. doi:10.1111/j.1365-2958.2009.06783.x

17. Akhter Y, Ehebauer MT, Mukhopadhyay S, Hasnain SE. The PE/PPE multigene family codes for virulence factors and is a possible source of mycobacterial antigenic variation: perhaps more? *Biochimie* (2012) **94**:110–6. doi:10.1016/j.biochi.2011.09.026

18. Vordermeier HM, Hewinson RG, Wilkinson RJ, Wilkinson KA, Gideon HP, Young DB, et al. Conserved immune recognition hierarchy of mycobacterial PE/PPE proteins during infection in natural hosts. *PLoS One* (2012) **7**:e40890. doi:10.1371/journal.pone.0040890

19. Abdallah AM, Gey van Pittius NC, Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CM, et al. Type VII secretion – mycobacteria show the way. *Nat Rev Microbiol* (2007) **5**:883–91. doi:10.1038/nrmicro1773

20. Stoop EJ, Bitter W, van der Sar AM. Tubercle bacilli rely on a type VII army for pathogenicity. *Trends Microbiol* (2012) **20**:477–84. doi:10.1016/j.tim.2012.07.001

21. Simeone R, Bottai D, Brosch R. ESX/type VII secretion systems and their role in host-pathogen interaction. *Curr Opin Microbiol* (2009) **12**:4–10. doi:10.1016/j.mib.2008.11.003

22. Brodin P, Rosenkrands I, Andersen P, Cole ST, Brosch R. ESAT-6 proteins: protective antigens and virulence factors? *Trends Microbiol* (2004) **12**:500–8. doi:10.1016/j.tim.2004.09.007

23. Houben EN, Bestebroer J, Ummels R, Wilson L, Piersma SR, Jimenez CR, et al. Composition of the type VII secretion system membrane complex. *Mol Microbiol* (2012) **86**:472–84. doi:10.1111/j.1365-2958.2012.08206.x

24. Jones GJ, Gordon SV, Hewinson RG, Vordermeier HM. Screening of predicted secreted antigens from *Mycobacterium bovis* reveals the immunodominance of the ESAT-6 protein family. *Infect Immun* (2010) **78**:1326–32. doi:10.1128/IAI.01246-09

25. Arlehamn CS, Sidney J, Henderson R, Greenbaum JA, James EA, Moutaftsi M, et al. Dissecting mechanisms of immunodominance to the common tuberculosis antigens ESAT-6, CFP10, Rv2031c (hspX), Rv2654c (TB7.7), and Rv1038c (EsxJ). *J Immunol* (2012) **188**:5020–31. doi:10.4049/jimmunol.1103556

26. Aagaard C, Brock I, Olsen A, Ottenhoff TH, Weldingh K, Andersen P. Mapping immune reactivity toward Rv2653 and Rv2654: two novel low-molecular-mass antigens found specifically in the *Mycobacterium tuberculosis* complex. *J Infect Dis* (2004) **189**:812–9. doi:10.1086/381679

27. Brock I, Weldingh K, Leyten EM, Arend SM, Ravn P, Andersen P. Specific T-cell epitopes for immunoassay-based diagnosis of *Mycobacterium tuberculosis* infection. *J Clin Microbiol* (2004) **42**:2379–87. doi:10.1128/JCM.42.6.2379-2387.2004

28. McEvoy CR, Cloete R, Muller B, Schurch AC, van Helden PD, Gagneux S, et al. Comparative analysis of *Mycobacterium tuberculosis* pe and ppe genes reveals high sequence variation and an apparent absence of selective constraints. *PLoS One* (2012) **7**:e30593. doi:10.1371/journal.pone.0030593

29. Blythe MJ, Zhang Q, Vaughan K, de Castro R Jr, Salimi N, Bui HH, et al. An analysis of the epitope knowledge related to Mycobacteria. *Immunome Res* (2007) **3**:10. doi:10.1186/1745-7580-3-10

30. Uplekar S, Heym B, Friocourt V, Rougemont J, Cole ST. Comparative genomics of Esx genes from clinical isolates of *Mycobacterium tuberculosis* provides evidence for gene conversion and epitope variation. *Infect Immun* (2011) **79**:4042–9. doi:10.1128/IAI.05344-11

31. Sayes F, Sun L, Di Luca M, Simeone R, Degaiffier N, Fiette L, et al. Strong immunogenicity and cross-reactivity of *Mycobacterium tuberculosis* ESX-5 type VII secretion: encoded PE-PPE proteins predicts vaccine potential. *Cell Host Microbe* (2012) **11**:352–63. doi:10.1016/j.chom.2012.03.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.