International Journal of Infectious Diseases 56 (2017) 155–166

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Review

Immune recognition surface construction of *Mycobacterium tuberculosis* epitope-specific antibody responses in tuberculosis patients identified by peptide microarrays

Davide Valentini[a,b], Martin Rao[b], Giovanni Ferrara[c,d], Marc Perkins[e], Ernest Dodoo[b], Alimuddin Zumla[f], Markus Maeurer[a,b,*]

[a] *Centre for Allogeneic Stem Cell Transplantation (CAST), Karolinska University Hospital Huddinge, Stockholm, Sweden*
[b] *Division of Therapeutic Immunology (TIM), Department of Laboratory Medicine (LABMED), Karolinska Institutet, Stockholm, Sweden*
[c] *Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden*
[d] *Department of Respiratory Medicine and Allergy, Karolinska University Hospital, Solna, Sweden*
[e] *FIND, Geneva, Switzerland*
[f] *Centre for Clinical Microbiology, Division of Infection and Immunity, University College London, and NIHR Biomedical Research Centre, UCL Hospitals NHS Foundation Trust, London, UK*

ARTICLE INFO

Article history:
Received 1 November 2016
Accepted 14 January 2017
Corresponding Editor: Eskild Petersen, Aarhus, Denmark

Keywords:
Tuberculosis
Mycobacterium tuberculosis
Peptide microarray
Immune recognition surfaces
Humoral immune response

SUMMARY

Background: Understanding of humoral immune responses in tuberculosis (TB) is gaining interest, since B-cells and antibodies may be important in diagnosis as well as protective immune responses. High-content peptide microarrays (HCPM) are highly precise and reliable for gauging specific antibody responses to pathogens, as well as autoantigens.
Methods: An HCPM comprising epitopes spanning 154 proteins of *Mycobacterium tuberculosis* was used to gauge specific IgG antibody responses in sera of TB patients from Africa and South America. Open source software for general access to this method is provided.
Results: The IgG response pattern of TB patients differs from that of healthy individuals, with the molecular complexity of the antigens influencing the strength of recognition. South American individuals with or without TB exhibited a generally stronger serum IgG response to the tested *M. tuberculosis* antigens compared to their African counterparts. Individual *M. tuberculosis* peptide targets were defined, segregating patients with TB from Africa versus those from South America.
Conclusions: These data reveal the heterogeneity of epitope-dependent humoral immune responses in TB patients, partly due to geographical setting. These findings expose a new avenue for mining clinically meaningful vaccine targets, diagnostic tools, and the development of immunotherapeutics in TB disease management or prevention.

Contents

* Corresponding author.
E-mail address: markus.maeurer@ki.se (M. Maeurer).

http://dx.doi.org/10.1016/j.ijid.2017.01.015

Introduction

The quest to discover novel antigenic targets for the differential diagnosis of active and latent tuberculosis (TB) presents a challenging task. Although efforts are manifold and performed on a global scale, biological markers derived from *Mycobacterium tuberculosis* that can precisely distinguish between the various stages of *M. tuberculosis* infection and disease remain elusive.[1] Several field-based studies published in recent years have indeed provided valuable information while enriching the knowledge base pertinent to TB immunodiagnostics and clinically relevant immune responses.[2,3] Nevertheless, predominant factors, such as geographical setting of the study locations, socio-economic status of the study populations, and risk and rate of exposure to various strains of *M. tuberculosis* therein, in addition to the intricate biology of the pathogen itself, inevitably represent confounding factors.[4] In this regard, multiplatform analysis of immune responses in patients with TB, as well as individuals harbouring latent TB infection (LTBI), has initiated great interest due to the wealth of information made accessible to the global TB community.[1]

Identifying specific *M. tuberculosis*-derived antigenic determinants of adaptive immune responses (orchestrated by T- and B-cells), especially in individuals with LTBI, is clinically relevant since it will impact on patient treatment regimens. Due to the immense volume of data that is subsequently generated, stringent quality control measures are required in order to obtain a more true-to-life picture of pertinent immune responses.

M. tuberculosis epitope mining of 12-mer linear peptides derived from *M. tuberculosis* antigens, displayed on a microarray slide, has been reported previously.[5] This technology is based on the recognition and binding of antigen-specific human serum IgG to the respective cognate epitope. Differential antigen recognition patterns were observed between patients with pulmonary TB and healthy individuals from Armenia.[5] This method was further validated using soluble human leukocyte antigen (HLA) class II molecules, which could form complexes after binding to their respective peptides (derived from 61 different *M. tuberculosis* proteins) on the slide, thus providing additional information on the allelic restrictions for HLA-based presentation of immunogenic *M. tuberculosis* epitopes to CD4+ T-cells.[6] Three *M. tuberculosis* proteins first discovered to be immunologically relevant using this platform were evaluated among Honduran TB patients and healthcare workers exposed to *M. tuberculosis*,[7] as well as Belarussian patients with pulmonary TB.[8] The present study group is currently evaluating all three proteins as vaccine candidates in preclinical studies.

In this study, the serum antibody responses of patients with pulmonary TB from two different geographical regions with a high disease burden (Africa and South America) were evaluated using the peptide microarray platform. The recognition of specific *M. tuberculosis* epitopes in patients with active TB (TB+), as well as those without clinical disease (TB−), was assessed to generate an overall humoral immune response landscape. This information allowed the exact epitopes of immunogenic *M. tuberculosis* antigens that elicit measurable immune responses to be uncovered for the first time, and furthermore, the pattern of this response in health and disease, using immune recognition surfaces, by describing the chemical composition of antibody recognition independent of the biological nature of the respective *M. tuberculosis* antigen.

Materials and methods

Study subjects

Serum samples were provided by the Public Health Research Institute (PHRI), New Jersey, USA, and stored at −70 °C. Volunteers (TB patients and healthy individuals) classified under two geographically distinct cohorts were involved in this study: Africa ($n = 228$) and South America ($n = 139$). After matching by age and sex, only 120 patients (30 individuals per cohort) were included in the final analysis. The cohort description is provided in Table 1.

Slide production, scanning, and analysis

The peptide arrays were custom-manufactured by JPT (Berlin, Germany) as reported previously.[9] Slides consist of three identical sub-arrays, each with 6720 spots arranged in 16 blocks of 420 spots. Of the 6720 spots, 485 are negative 'empty' control spots, 271 are positive control spots (144 peptide controls, as described in Ngo et al.[10]), and 5694 are unique peptides generated from 154 *M. tuberculosis* proteins (listed in the **Supplementary Material**, Table S3) as 15-mers overlapping by five amino acid residues.

Sera were diluted 1:100 in 300 μl buffer (phosphate buffered saline (PBS), 3% foetal calf serum (FCS), 0.5% Tween 80; Sigma Aldrich, St Louis, MO, USA) and added to microarray chips for 16 h of incubation in a humid chamber at +4 °C. The slides were then washed with buffer twice and sterile distilled water three times. This was followed by secondary incubation with 300 μl diluted (1:500) Cy5-labelled mouse anti-human IgG monoclonal antibody IgG secondary antibody (catalogue number 6561-100; Abcam, UK) for 1 h at room temperature (RT) and then washing as before. The slides were then spun dry in a slide centrifuge (DJB Labcare, Newport Pagnell, UK).

Table 1
Description of the study cohorts from Africa and South America.

Continent	TB status	Number	Mean age (SD)	M:F ratio
Africa	Positive	89	37.6 (±15.2)	2.0
	Negative	140	40.2 (±16.2)	0.8
Total		228	39.2 (±15.8)	1.1
South America	Positive	72	33.9 (±12.7)	1.2
	Negative	67	41.5 (±15.6)	0.6
Total		139	37.6 (±14.6)	0.9

TB, tuberculosis; SD, standard deviation; M, male; F, female.

Each slide was scanned with the GenePix 4000 B microarray scanner (Molecular Devices) at a wavelength of 635 nm, and the resulting images were saved in high-resolution format. Image analysis was performed using GenePix 7.0 image analysis software (Molecular Devices) and the GenePix Array List (GAL) file with the peptide location information for each spot as supplied by the manufacturer (JPT, Berlin, Germany). The detectable spots not internally uniform were flagged as 'bad' (i.e., unreliable). This is efficiently estimated using the following criteria (described in detail in Perez-Bercoff et al.[11]): ([F635 Mean] > (1.5*[F635 Median])) AND ([F635 Median] > 40), which identifies the spots with a mean foreground value different from the spots exhibiting median fluorescence intensity values. The spots were visually inspected after flagging and corrected as necessary. The results for each sub-array were saved separately.

Data mining and statistical analyses

Software

All of the methods described here were implemented using the open source R language and packages from Bioconductor project.[12,13]

Data pre-processing

GenePix results files were read and an R dataset was created. Digital images of background and foreground intensities were produced for every sub-array to detect and exclude potential signal artefacts. Signals were corrected for background noise by computing the index value (log2 foreground/background[14]). All spots or areas not representing a high quality signal were not analysed. Further quality controls included scatter plots (index vs. log2 background) for each slide to remove outliers and abnormal values and also scatter plots (average index vs. average log2 background) for all slides in each group to address the efficacy of the negative and positive controls. A detailed control of flag distribution proportions (−100, −75, −50, 0) was done for the whole dataset and separately for each group analysed. Potential false-positives were identified[14]: all spots showing a response on the buffer plus secondary antibodies slides were labelled and removed from the analysis.

Normalization

The normalization process was performed separately for each geographical group (Africa and South America) using the linear model described previously[9]: $I = slide_i + subarray_j + block_k + \varepsilon$, where I is the measured intensity; $slide_i$ is the slide effect, i.e., the portion of the signal intensity due to the presence of the spot on the slide i; $subarray_j$ is the sub-array effect, i.e., the portion of the intensity due to the presence of the spot at one of the sub-arrays in the slide; $block_k$ is the block effect or the effect on the intensity due to the position of the spot in one of the blocks in the sub-array; ε is the residual, composed from biological interaction and slide and sub-array interaction ($slide_i * subarray_j$). Data were fitted into the linear model and the estimated slide, sub-array, and block effects removed. The quality of the normalization was assessed by visual inspection of the normalized data plot in all of the study groups.

Differential recognition

The differential recognition was performed using the empirical Bayes moderated t-statistics function contained in the R/Bioconductor package limma.[15] Peptide p-values were adjusted using the Benjamini–Hochberg method (control of FDR (false discovery rate)) available in the same package. For the African cohort, the data first underwent a filtering procedure to reduce the heterogeneity within groups. Peptides with very high coefficient of variation (CV) within groups or very low CV between groups did not enter the analysis. Group outlier slides (top or bottom 5% of the cohort) were also removed from the differential recognition analysis.

Pept3D: surface regression and three-dimensional visualization

The normalized data were visualized on a three-dimensional (3D) surface according to different methods (Figure 1). The base functions are BP3D, Pos3D, and additionally Diff3D, and provide base coordinates for surface plots based on (1) amino acid characteristics (i.e. bulkiness/polarity plot, Figure 1a); (2) the peptide response along ranked proteins (i.e., proteins plot, Figure 1b); (3) the surface of the difference in the immune response between either (1) or (2), respectively. For each of these functions, an R script was produced and made available in the package Pept3D (downloadable at http://ki.se/en/people/davval).

BP3D. For each of the peptides included in the analysis, two values are computed:

$$\text{Relative peptide bulkiness (RPB)} = \frac{\sum_{i=1}^{n} \beta_i}{n\beta_{MAX}}$$

where β_i is the bulkiness value of the i^{th} amino acid in the peptide sequence,[16] β_{MAX} is the maximum theoretical bulkiness value (21.670, from tryptophan, in the bulkiness amino acid scale by Zimmerman et al.[16]), and n is the number of amino acids in the peptidic sequence. The bulkiness values for each amino acid are reported in the **Supplementary Material** (Table S1).

$$\text{Relative peptide polarity (RPP)} = \frac{\sum_{i=1}^{n} \pi_i}{n\pi_{MAX}}$$

where π_i is the polarity value of the i^{th} amino acid in the sequence (the polarity scale of Grantham[17] was chosen), π_{MAX} is the

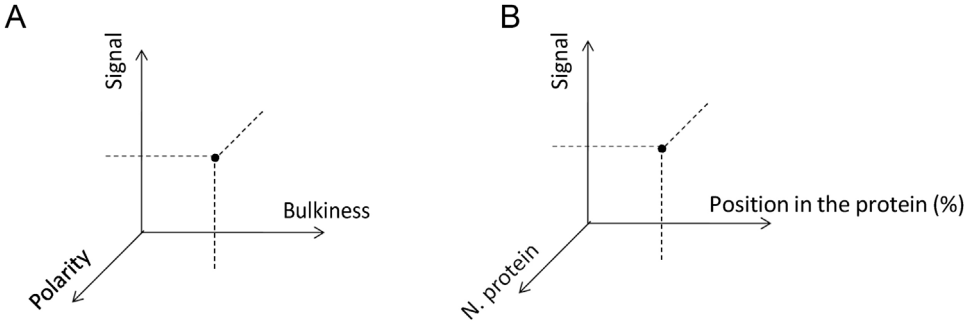

Figure 1. Visualization of the 3D plotting methods for (a) the bulkiness–polarity plot, and (b) the protein plot.

Table 2
Eight protein antigens of *Mycobacterium tuberculosis* found to be differentially recognized by serum antibodies from African and South American TB patients.

Peptide	*M. tuberculosis* protein (Rv number)	Remarks	Ref.
GAIPGGWWLTFGQIL	PPE32 (Rv1808)	Gene expression induced by restricted Mg^{2+} bioavailability during in vitro nutrient starvation	30
NKPVLVDFWATWCGP	Thioredoxin C, TrxC (Rv3914)	Biologically active sulphide reductase involved in hydroperoxide, dinitrobenzene, and ROI degradation	41,42,47
MFSGFDPWLPSLGNP	PPE2 (Rv0256c)	A secreted protein that may block nitric oxide production by activated macrophages	27
VYVAQKRKISDGDKL	RNA polymerase Beta subunit (RpoB) (Rv0667)	Mutant forms of this enzyme are associated with resistance to rifampicin	48
GGLFGIGGAGGGCGS	PE-PGRS41 (Rv2396)	Linked to PhoP regulation due to stark down-regulation of Rv2396 gene expression in PhoP-deficient *M. tuberculosis* H37Rv, and very strong binding site in chromosomal vicinity of Rv2396	30,32
IPIPGLPDIPGLPDF	EspE (Rv3864)	Associated with ESX-1 secretion system; localized in *Mycobacterium marinum* and *M. tuberculosis* H37Rv cell wall capsule; strongly recognized in sera of patients with cavitary TB	35,36
TRYSSAYELEGAVKR	TB43 (Rv2780)	43-kDa cell-bound protein first identified as a pyridine nucleotide transhydrogenase; later recognized as a 40-kDa mature protein with alanine dehydrogenase activity crucial for peptidoglycan turnover	44,45
ARTISEAGQAMASTE	ESAT-6 (Rv3875)	Important secreted antigen with pore-forming characteristics found in *M. tuberculosis* and other pathogenic mycobacteria; used in immunodiagnostics for the detection of latent TB infection	37

ROI, reactive oxygen intermediate.

maximum theoretical polarity value (13.00, from aspartic acid), and n is the number of amino acids in the peptide sequence (values are provided in the **Supplementary Material**, Table S2).

The values of peptide relative polarity, relative bulkiness, and normalized signal intensity are then plotted on a 3D graph as x, z, and y coordinates, respectively, using a non-parametric smoothing regression technique and surface plot provided by the R function *sm.regression*, included in the package *sm*.[18] The drawn surface is coloured according to the values of the y-axis (signal intensities).

Pos3D. Complete amino acid sequences of the source proteins for the peptides (based on Uniprot and NCBI protein identifiers/accession numbers) were retrieved using the R function *getSeq* in the package *annotate* in the Bioconductor project. Protein identifiers are usually provided by the slide manufacturer. The start and end positions of each peptide along the amino acid sequence of the corresponding protein are determined, followed by computation of the percentile of the protein sequence marking the centre point of the peptide. The source proteins are then ranked and numbered according to their biological function in relation to immunological significance (**Supplementary Material**, Table S3). The default ranking system is based on alphabetical ordering of the adopted protein reference code (i.e., accession number), but users can specify their ranking system of choice. Next, the peptide centre points (x), corresponding protein rank numbers (y), and the normalized signal intensities (z) are plotted in the 3D graph using the *sm.regression* function as for BP3D. In addition, the smoothed regression bi-dimensional curve for a specific selected protein can also be plotted (Plot2D function).

Diff3D. This additional function allows the difference in the immune response of two individuals (or groups) in two different clinical conditions, or one individual at two different time points, to be estimated and this information to be plotted according to the coordinates obtained from the BP3D and Pos3D methods described above. This is done by computing the relative peptide difference (RPD) between two vectors of peptide intensities, x and y:

$$\text{Relative peptide difference } (\text{RPD})_{x,y} = \frac{x^* - y^*}{y^*} \times 100$$

where x^* and y^* are computed as:

$$x^* = x + |\min(x, y)|$$
$$y^* = y + |\min(x, y)|$$

in order to always have positive input values for the relative differences computation. The original normalized values may in fact be negative, since they are calculated as residuals from the fit of the linear model described above. In addition to this, the computed RPDs may have extreme outliers, which, as experimentally verified, can decrease the readability of the plot by adding isolated spikes on the surface. The function thus includes an outlier detection and removal procedure, where the user can specify the borderline value for the detection through a k parameter (how many times the interquartile range has to be distant from the upper or lower quartile). The RPDs are then plotted according to the desired set of coordinates, computed by either BP3D or Pos3D.

Detection of IgG in patient sera using high-affinity human IgG ELISA

Streptavidin-coated 96-well plates (Thermo Fisher, MA, USA) were adsorbed with 1 μg/ml of biotinylated *M. tuberculosis* peptides prepared in PBS (Table 2). An 8-point serial dilution (1:2 ratio) of recombinant human IgG standard (Sigma Aldrich) was prepared from a 15 000 ng/ml stock solution in PBS, and each dilution was added in duplicate to the assay plate. The plates were incubated for 1 h at 37 °C and washed thoroughly with PBS–0.05% Tween 20 (PBST20) buffer, followed by blocking with assay buffer (PBST20 + 0.1% bovine serum albumin (BSA)) at RT for 1 h. After washing with PBST20 buffer, serum samples diluted at 1:100 in assay buffer were added to the assay plate and incubated at RT for 2 h. After washing the plate with PBST20 buffer, alkaline phosphate-conjugated anti-human IgG detection monoclonal antibody (Mabtech, Stockholm, Sweden) diluted at a ratio of 1:1000 in assay buffer was added to all wells, followed by incubation for 1 h at RT. The assay plate was washed with PBST20 buffer post-incubation, followed by the addition of the substrate solution (*p*-nitrophenyl phosphate tablets (PNPP; Thermo Fisher, MA, USA) dissolved in 1 × diethanolamine substrate buffer (Thermo Fisher, MA, USA)). After incubating the assay plate at RT in the dark for 20–30 min, the enzymatic reaction was stopped with 1 N NaOH, and the optical density measured at 405 nm using a VMax Kinetic ELISA Microplate Reader (Molecular Devices, CA, USA).

Results

Differential recognition analysis

Peptide microarrays revealed differential recognition of *M. tuberculosis* proteins by serum IgG in TB patients and healthy individuals from two regions comprising high-burden countries for TB, namely Africa and South America.[19] Based on initial screening with the peptide microarray platform, geographical location represented a major confounding factor, and the differences in peptide recognition between TB+ and TB− individuals were much more evident after data stratification by the geographical area. The peptide microarray differential recognition

analysis (TB+ (patients) vs. TB− (healthy individuals)) was performed separately for these two cohorts. Analysis of the African cohort identified six differentially recognized peptides (five highly recognized and one poorly recognized) (**Supplementary Material**, Table S4), representing four different proteins (three peptides from the same protein, PPE2 (proline-proline-glutamic acid)). Conversely, differential recognition in the South American cohort showed 18 differentially recognized peptides, plus four with a borderline adjusted *p*-value (12 highly recognized and 10 poorly recognized) (**Supplementary Material**, Table S5).

Pept3D immune recognition surfaces

Protein curves

The Pos3D curves for peptide-specific IgG responses spanning entire *M. tuberculosis* protein sequences in African TB+ and TB− individuals are shown in Figure 2 (Figure 2A and B). A common response pattern was found to exist between the TB+ and TB− groups, respectively; i.e., peptides belonging to the first half of the ranked proteins (vicinity of PPE45–PPE55) elicited more robust antibody responses than the second half (**Supplementary Material**, Table S3). Although not strikingly visible, some differences were present and detectable in the corresponding Diff3D curve of the relative differences between the two groups. A marked difference in response was detected at the very beginning of the

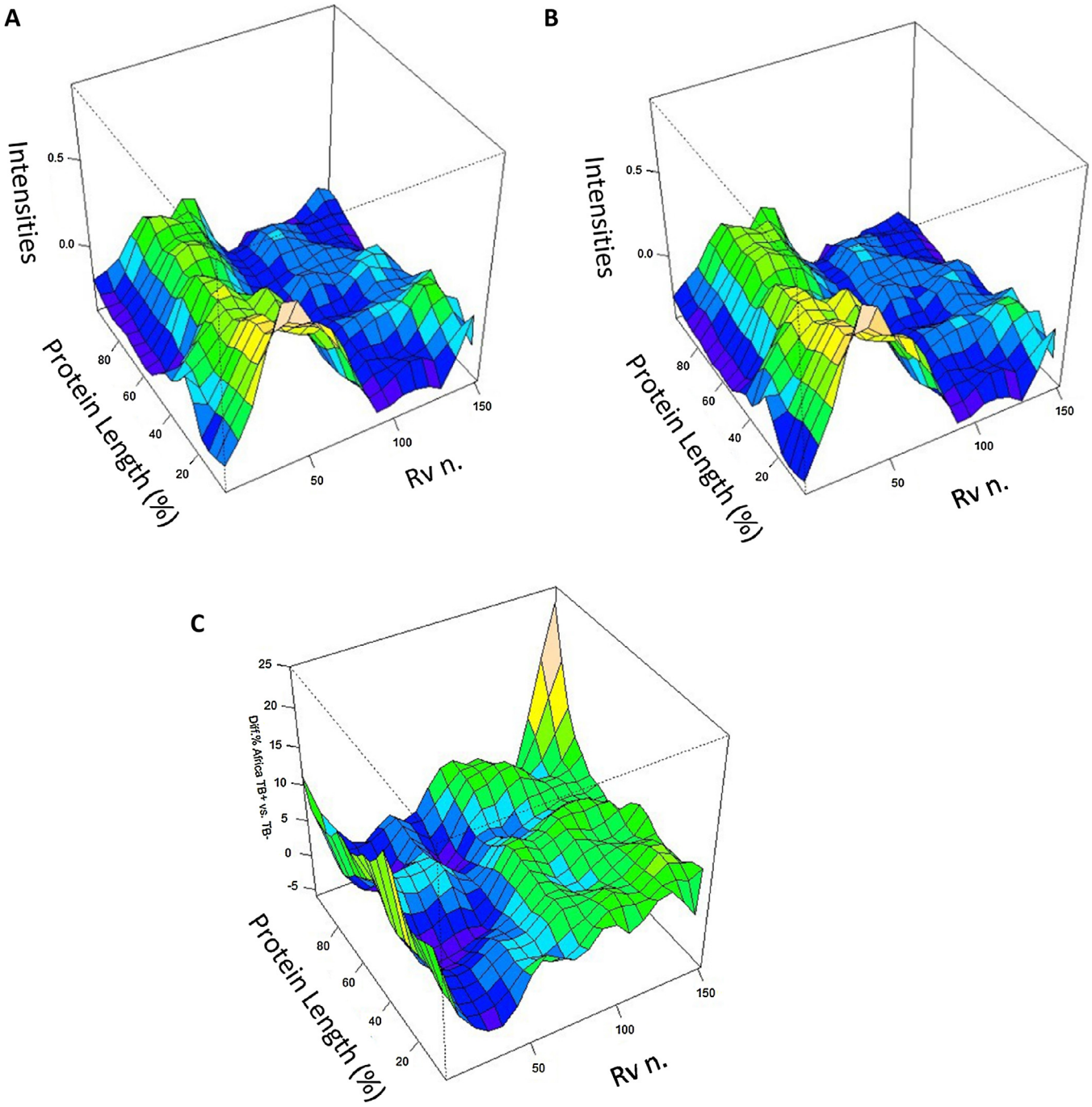

Figure 2. Immune response surface for peptides across proteins in African TB+ and TB− individuals (a and b), and surface analyses recognition differences (c).

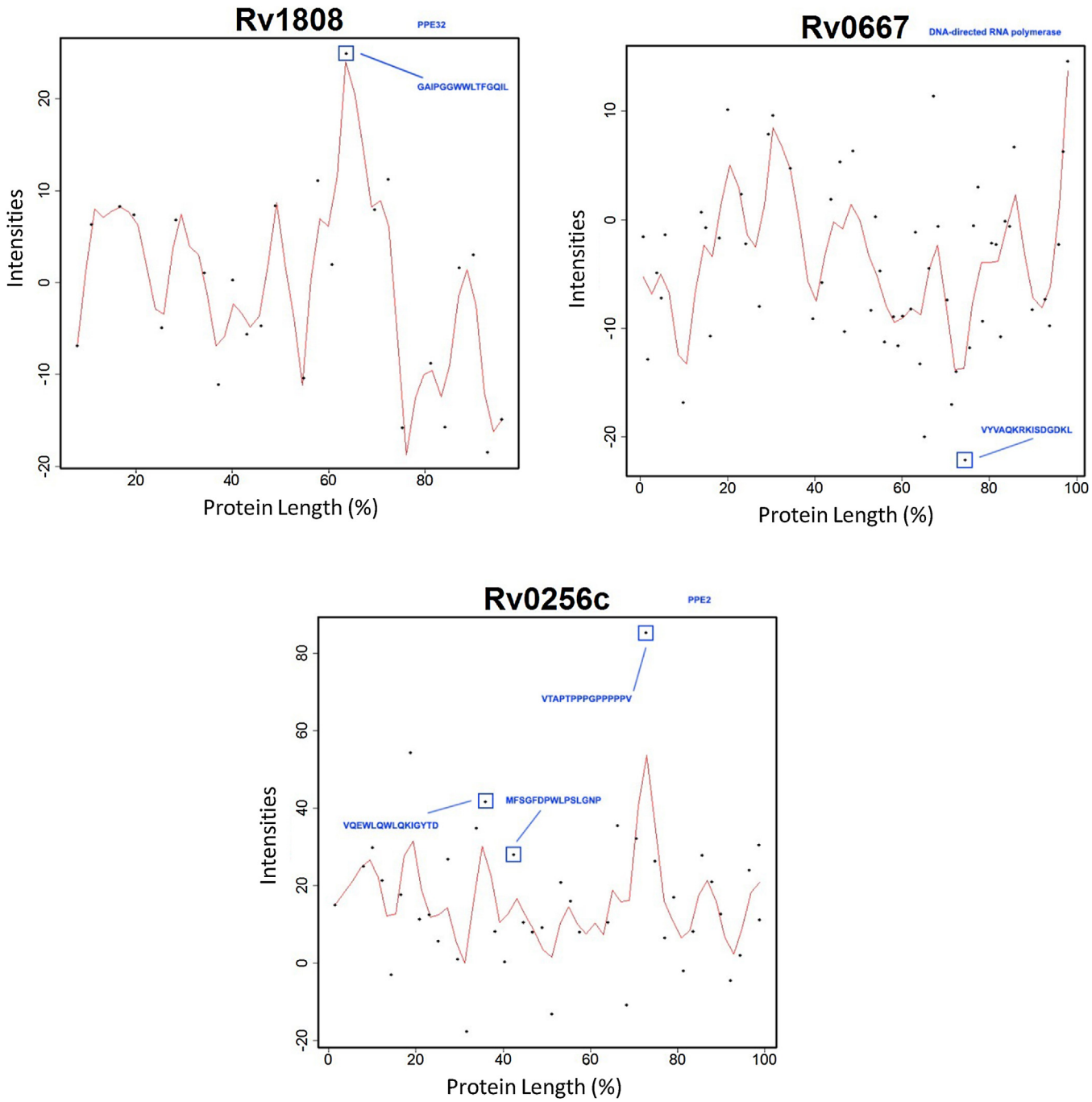

Figure 3. Relative differences in peptide responses (dots) and smoothed regression curve (plot 2D) for three selected TB proteins in the African cohort. Resulting peptides from the TB+ vs. TB− comparison are highlighted.

ranked protein list, corresponding to the first set of PPE proteins. Other 'hotspots' of recognition were present in the second part of the ranked protein list, particularly.

As a paradigm, two-dimensional (2D) curves were created for some of the proteins relative to the peptides identified with the differential recognition analysis (Figure 3). The peptide values were reported alongside the smoothed regression curve, thus highlighting the significantly differentially recognized peptides. All of the differentially recognized peptides with limma analysis corresponded to the highest (or lowest, when poorly recognized) peaks of the curve illustrated using this model.

The immune recognition surfaces for the South American cohort reaffirmed that the IgG response profiles of TB+ and TB− individuals, respectively, between the two cohorts were generally similar, although the South American individuals appeared to mount an overall more intense response compared to their African counterpart groups (Figure 4A and B). Differential peptide recognition between South American TB patients and healthy individuals revealed more robust responses by the former group to the PPE, as well as the other *M. tuberculosis* peptides tested, compared to the African TB patients (Figure 4C).

In addition, 2D plots of the peptides identified with the differential recognition analysis in the South American cohort emerged at the highest peaks in the curves (Figure 5).

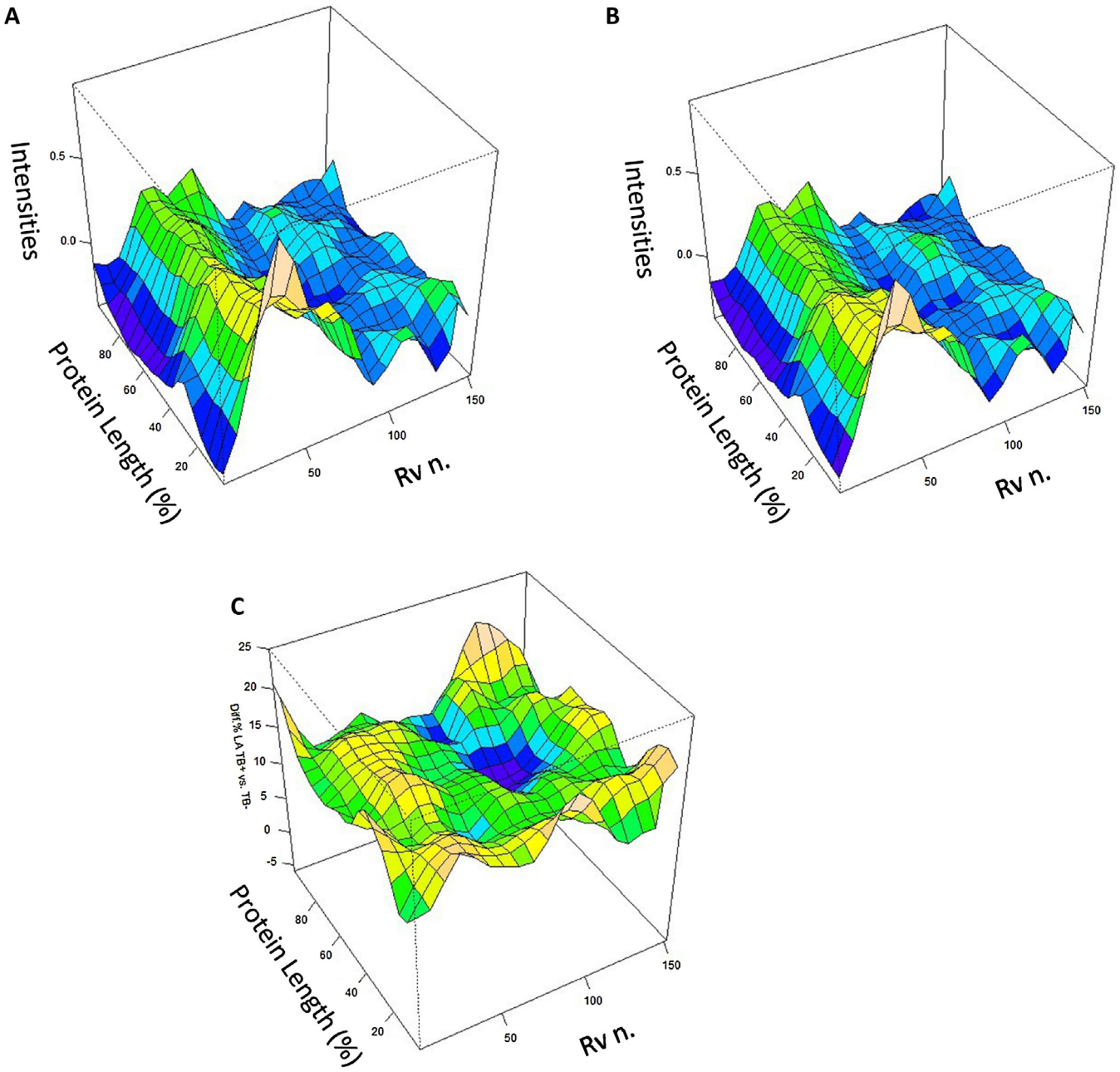

Figure 4. Immune response surface for peptides across proteins in South American TB+ and TB− individuals (a and b), and surface of the differences (c).

Bulkiness–polarity 3D curves

The humoral immune response landscape based on biochemical characteristics of the peptides (protein bulkiness and polarity dependent on amino acid composition; **Supplementary Material**, Tables S1 and S2) was constructed for both the African and South American cohort (Figures 6 and 7). This visualization clearly identified a characteristic shape: among TB+ and TB− individuals in both geographical groups, the immune response landscape showed that with increasing protein bulkiness (and to some extent with decreasing polarity – thus increasing molecular complexity), the response becomes less pronounced (Figure 6A and B, Figure 7A and B). Conversely, the immune response landscape of TB+ patients compared to TB− individuals in both cohorts (Fig. 6C and 7C) indicated enhanced recognition with increasing molecular complexity and diminishing polarity. Thus, TB patients are more likely to mount a stronger response to more complex protein antigens.

Antigen-specific serum IgG ELISA

Four differentially recognized peptides per cohort (one peptide per protein, in total eight peptides representing eight proteins) were selected for high-affinity human IgG ELISA, which detects the presence of antigen-specific IgG antibodies in serum (Table 2). Contrary to expectations, no significant differences in recognition of the eight *M. tuberculosis* peptides were found between patients with TB and healthy individuals (**Supplementary Material**, Fig. S1). However, it was striking to notice that antibody recognition of PE-PGRS41 by TB patients from South America was more pronounced than that of ESAT-6 (early-secreted antigenic target 6 kDa) (**Supplementary Material**, Fig. S2). In the African cohort, the strongest serum IgG reactivity was found to be directed against rpoB (RNA polymerase β-subunit), between patients with TB and healthy individuals alike. Other antigens

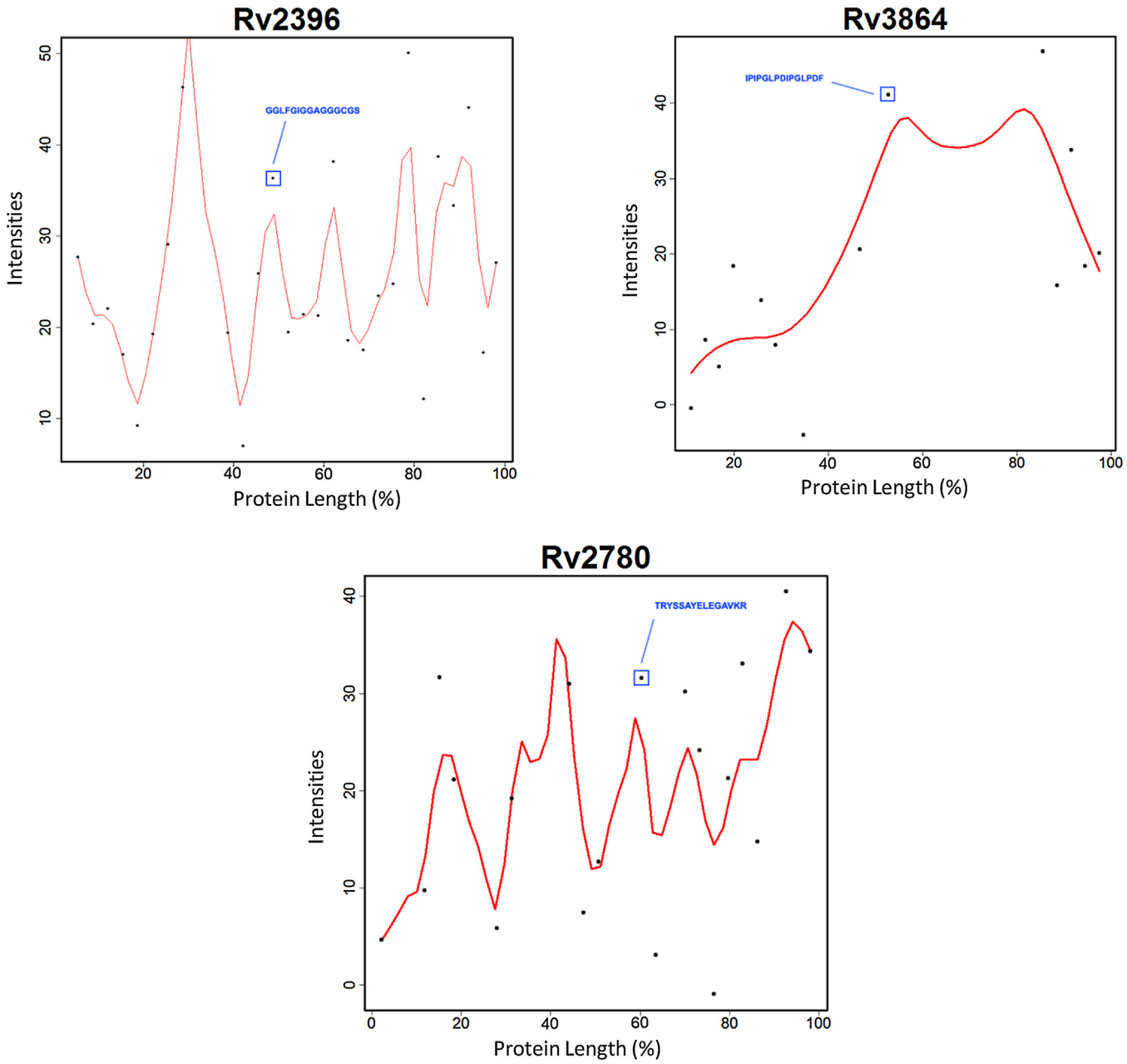

Figure 5. Relative differences in peptide responses (dots) and smoothed regression curve (plot 2D) for three selected TB proteins in the South American cohort. Resulting peptides from the comparison TB+ vs. TB− are highlighted.

tested for in this cohort were PPE32, PPE2, and trxC – all with implications in mycobacterial stress resistance.

Discussion

Previously, the detection and quantification of antigen-specific antibody responses in a given clinical condition was obtainable only via single-plex ELISA (antibodies specific for only one antigen), which is both cumbersome and time-consuming. Peptide microarrays enable the assessment of the entire expanse of an individual's humoral immune response to a given pathogen, or mutated self-antigens (such as in cancer) on a single chip. Additionally, results from peptide microarray studies reveal the various epitopes that are most commonly recognized by individuals as a means to provide 'hotspots' of immune recognition and epitope focus in a specified clinical context. This warrants faster data turnover, as well as accurate fishing of relevant antigenic targets for further validation. The necessity to provide open access to source software tools for applying the methods used in this study is evident.

The curves produced using the Pos3D and Diff3D methods have been introduced briefly in previous studies.[20–22] Detailed descriptions of these methods are provided here for the first time, and the 2D representation of data is introduced. In addition, a novel method based on amino acid biochemical characteristics (bulkiness and polarity), BP3D, is presented. These biochemical characteristics of the *M. tuberculosis* peptides shown in the 3D and 2D curves directly influence the likelihood and frequency of recognition by serum antibodies – reflective of antigen processing and presentation in the host. The proposed methods therefore give a 'photographic snapshot' of the immune response of an individual (or groups of individuals), as well as the variation between tested

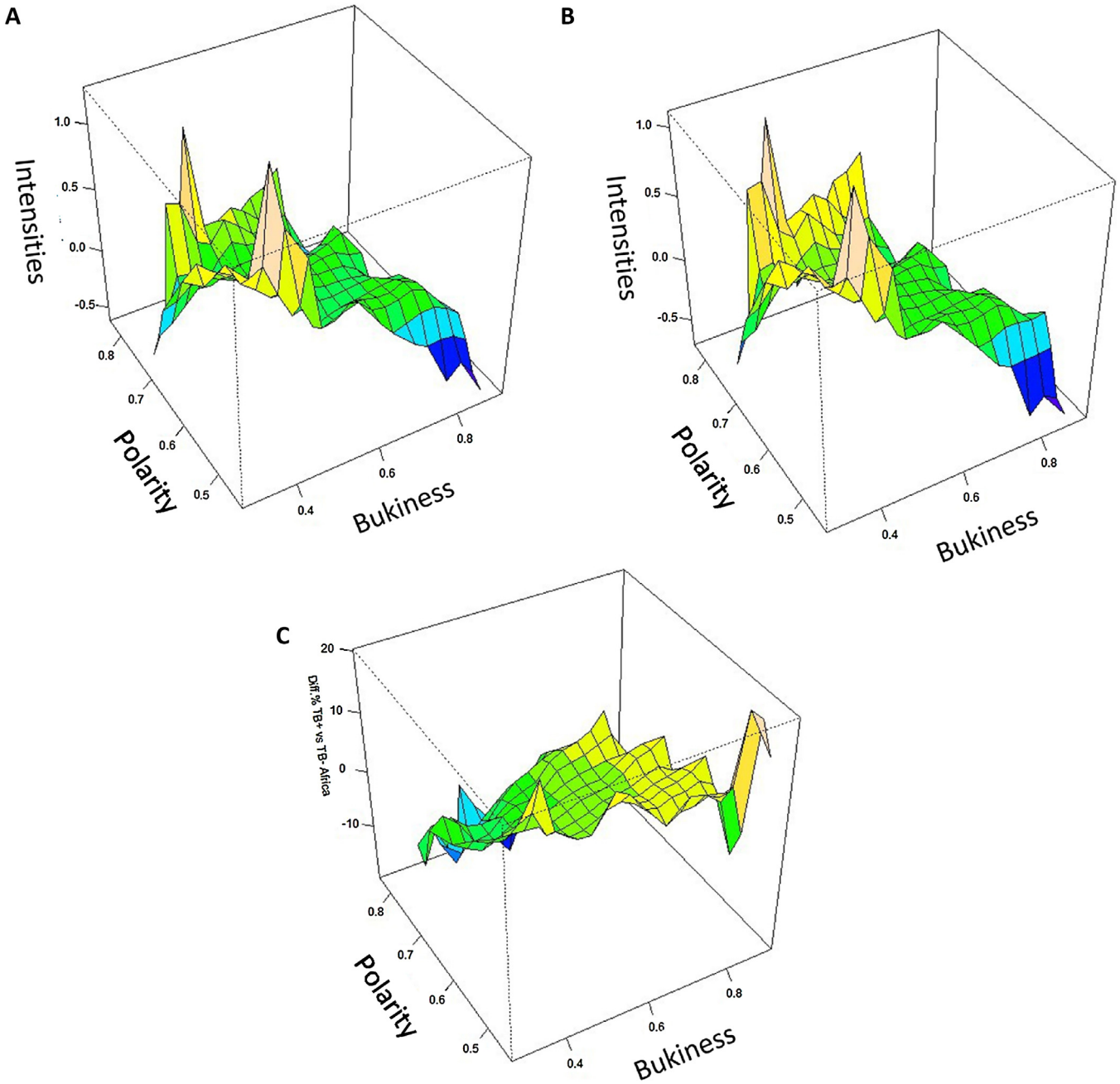

Figure 6. Immune response surface for peptide biochemical characteristics (bulkiness vs. polarity) in African TB+ and TB− individuals (a and b), and surface of the differences (c).

conditions (in this case, TB+ and TB− individuals, although the application of these methods stretch far and beyond). In addition, they can aid better visualization of the results obtained with other peptide analysis techniques for a better appreciation of the immune response and heterogeneity therein. This method captured the underlying biology reflected in humoral immune responses to *M. tuberculosis* antigens in two different continents, while acknowledging that the results observed are jointly influenced by a high number of factors, i.e., genetic differences, environmental pressure, socio-economic status, and co-morbidities, to name a few.

To the best of the authors' knowledge, this is the first time that 'peptide-level' information has been combined with 'protein-level' information to allow an insight into humoral immune responses not only to isolated sub-protein spots (single epitope), but across the entire protein molecule. Although the 3D structure of an antigen is important for its activity and recognition by the immune system, visualization of the peptides within a protein that evoke specific antibody responses in various clinical scenarios promises a wealth of information for vaccine development and diagnostics. In addition, the peptide microarray platform allows the detection of molecules that undergo conformational changes specific to disease conditions such as in prion disease, as well as mutated neoantigens in cancer.[23,24] Moreover, for the first time, a different view of peptide microarray results relative to their biochemical characteristics is proposed. The amino acid sequences in antigenic epitopes dictate their polarity, thus affecting their processing and presentation to T-cells, as well as binding to antibodies. For example, the presence of glutamine and glutamic acid in epitopes reduces their affinity for antibody binding, while an abundance of tyrosine and

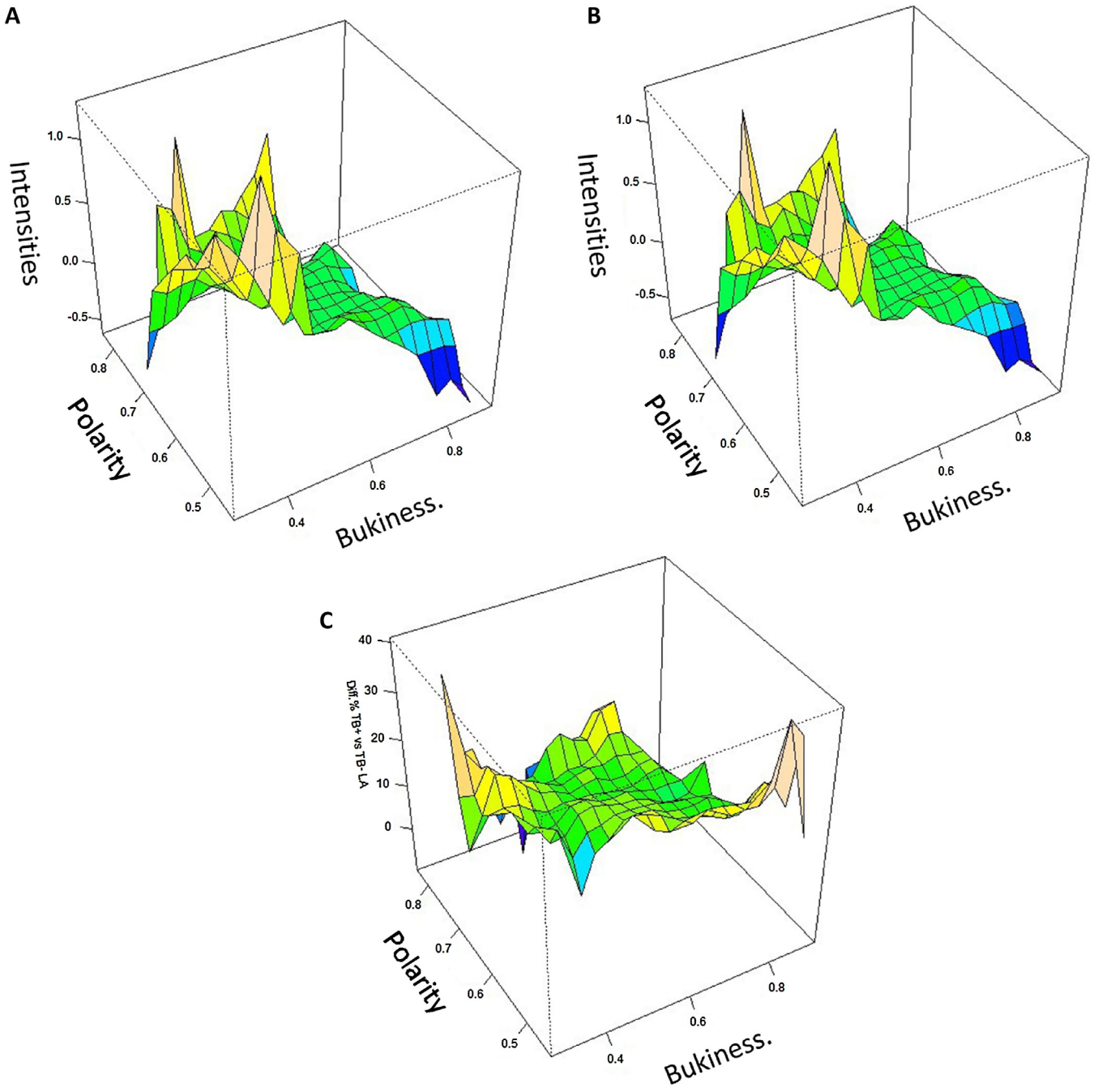

Figure 7. Immune response surface for peptide biochemical characteristics (bulkiness vs. polarity) in South American TB+ and TB− individuals (a and b), and surface of the differences (c).

tryptophan have the reverse effect.[25] Furthermore, epitopes with a higher antibody binding ratio are more likely to activate the major histocompatibility complex (MHC)-I and MHC-II pathways, allowing concerted T- and B-cell responses.[25] Therefore, the polarity of antigens may potentially influence the clinical relevance of ensuing adaptive immune responses in the host.

A summary of the eight *M. tuberculosis* proteins that were differentially recognized by TB patients in the African and South American cohorts is provided in Table 2. The antibody responses directed against these *M. tuberculosis* proteins are of important clinical significance, since they may protect against or perpetrate pathology in TB. In terms of mycobacterial physiology, *M. tuberculosis* H37Rv constitutively up-regulates PPE2 gene expression when subjected to various physicochemical treatments.[26] Also, murine macrophages infected with PPE2-deficient *M. tuberculosis* CDC1551 were able to generate more nitric oxide compared to their wild-type counterparts.[27] Gene expression studies with *M. tuberculosis* H37Rv revealed that PPE32 was highly expressed upon treatment with hydrogen peroxide for 40 min, as well as in the stationary phase of growth.[28] In addition, sodium dodecyl sulphate treatment of *M. tuberculosis* H37Rv also induces PPE32 expression.[29] PPE32 expression was up-regulated in PhoPQ-deficient *M. tuberculosis* H37Rv during in vitro growth in medium with limited Mg^{2+} supply.[30] Therefore, current evidence suggests that the PPE proteins may play a role in the stress response and intracellular survival of *M. tuberculosis* within macrophages, although this hypothesis requires further investigation.[26]

Proteins belonging to the polymorphic CG-repetitive sequences (PGRS) family have been used as beacons for molecular epidemiological studies performed on clinical isolates of *M. tuberculosis*.[31] PE_PGRS41 gene expression has been linked to regulation by the PhoPQ system, thus pathogenicity,[30,32] and shares 67% identity with PE_PGRS81 across 98 amino acids at the N-terminus.[33] Importantly, PE_PGRS81 (Rv1759c) has been associated with the ability to bind fibronectin,[34] and was first discovered with an antibody that recognizes the *Mycobacterium bovis* antigen 85 complex, the best-described mycobacterial fibronectin-binding protein.[33]

The mycobacterial capsular protein EspE has been reported to facilitate *M. tuberculosis* interaction with the host macrophage, and subsequent down-modulation of proinflammatory cytokine release.[35] EspE was also found to be closely associated with humoral immune responses in patients with active cavitary TB.[36] In addition, EspE is associated with the Esx-1 protein transport system of *M. tuberculosis*, the same apparatus responsible for CFP-10 (culture filtrate protein 10 kDa) and ESAT-6 secretion.[37] As such, EspE could potentially be evaluated as an immunological target.

ESAT-6 is the most immunologically relevant and best characterized protein listed in Table 2, and also exhibits pore-forming properties in addition to being involved in granuloma formation in the host.[38] Since ESAT-6 is enriched for human T-cell epitopes and is not encoded by the bacille Calmette–Guérin (BCG) vaccine, it is used in standard LTBI immunodiagnosis.[37,39] In addition, several TB vaccine candidates incorporating full-length ESAT-6 are presently in various stages of clinical evaluation.[40]

This is the first study to present thioredoxin C (TrxC) as an immunologically relevant *M. tuberculosis* target, although its involvement in mycobacterial detoxification mechanisms has been reported previously.[41–43,47] This likewise applies to TB43, whose role in *M. tuberculosis* cell wall maintenance is already known,[44,45] and rpoB, which has long been used as the genetic marker for *M. tuberculosis* resistance to rifampicin.[46–48] Further studies are required for validation of these proteins as legitimate biomarkers. It is also acknowledged that a majority of the peptides observed in the 3D curves require detailed characterization in the context of host–pathogen interaction, as well as their role in the course of clinical TB. The description of biochemical properties in antibody recognition allow for data mining for non-*M. tuberculosis*-related proteins, yet exhibiting similar biochemical properties, i.e., the induction of heterologous immune responses.

In conclusion, the immune response landscape in a cohort offers the unique opportunity to obtain a global view of the differential recognition of antigenic peptides by antibodies. This general picture can be used as a guide to 'cherry-pick' nominal antigenic targets occurring within hotspots of recognition for further validation. The selective immunogenicity of some peptides over others, derived from the same protein, can be determined from the 3D curves presented in this study. This information does not only enrich our knowledge of the host–pathogen interaction, but can be equally instrumental in basic and clinical research, stretching from primary evaluation of peripheral immune responses to large-scale immune-monitoring during clinical trials and anti-TB therapy. Analysis of the humoral immune landscape allows the objective identification of complex antibody recognition patterns.

Funding

This study was supported by grants from Hjärtlungfonden (Swedish Heart and Lung Foundation), the FIND Foundation, Vinnova and Vetenskapsrådet (Swedish Research Council) to MM.

Conflict of interest

The authors report no conflicts of interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijid.2017.01.015.

References

1. Maertzdorf J, Kaufmann SH, Weiner 3rd. J3rd.. Toward a unified biosignature for tuberculosis. *Cold Spring Harb Perspect Med* 2014;**5**(1):a018531.
2. Tucci P, Gonzalez-Sapienza G, Marin M. Pathogen-derived biomarkers for active tuberculosis diagnosis. *Front Microbiol* 2014;**5**:549.
3. Nayak K, Jing L, Russell RM, Davies DH, Hermanson G, Molina DM, et al. Identification of novel *Mycobacterium tuberculosis* CD4 T- cell antigens via high throughput proteome screening. *Tuberculosis* 2015;**95**:275–87.
4. Marais BJ, Lonnroth K, Lawn SD, Migliori GB, Mwaba P, Glaziou P, et al. Tuberculosis comorbidity with communicable and non-communicable diseases: integrating health services and control efforts. *Lancet Infect Dis* 2013;**13**:436–48.
5. Gaseitsiwe S, Valentini D, Mahdavifar S, Magalhaes I, Hoft DF, Zerweck J, et al. Pattern recognition in pulmonary tuberculosis defined by high content peptide microarray chip analysis representing 61 proteins from *M tuberculosis*. *PLoS One* 2008;**3**:e3840.
6. Gaseitsiwe S, Valentini D, Mahdavifar S, Reilly M, Ehrnst A, Maeurer M. Peptide microarray-based identification of *Mycobacterium tuberculosis* epitope binding to HLA-DRB1*0101, DRB1*1501, and DRB1*0401. *Clin Vaccine Immunol* 2010;**17**:168–75.
7. Alvarez-Corrales N, Ahmed RK, Rodriguez CA, Balaji KN, Rivera R, Sompallae R, et al. Differential cellular recognition pattern to M: *tuberculosis* targets defined by IFN-gamma and IL-17 production in blood from TB+ patients from Honduras as compared to health care workers: TB and immune responses in patients from Honduras. *BMC Infect Dis* 2013;**13**:125.
8. Ahmed RK, Rohava Z, Balaji KN, Hoffner SE, Gaines H, Magalhaes I, et al. Pattern recognition and cellular immune responses to novel *Mycobacterium tuberculosis*-antigens in individuals from Belarus. *BMC Infect Dis* 2012;**12**:41.
9. Nahtman T, Jernberg A, Mahdavifar S, Zerweck J, Schutkowski M, Maeurer M, Reilly M. Validation of peptide epitope microarray experiments and extraction of quality data. *J Immunol Methods* 2007;**328**:1–13.
10. Ngo Y, Advani R, Valentini D, Gaseitsiwe S, Mahdavifar S, Maeurer M, Reilly M. Identification and testing of control peptides for antigen microarrays. *J Immunol Methods* 2009;**343**:68–78.
11. Perez-Bercoff L, Valentini D, Gaseitsiwe S, Mahdavifar S, Schutkowski M, Poiret T, et al. Whole CMV proteome pattern recognition analysis after HSCT identifies unique epitope targets associated with the CMV status. *PLoS One* 2014;**9**:e89648.
12. R Core Team. *R: a language and environment for statistical computing, R Foundation for Statistical Computing* 2014. Vienna, Austria.
13. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;**5**:R80.
14. Reilly M, Valentini D. Visualisation and pre-processing of peptide microarray data. *Methods Mol Biol* 2009;**570**:373–89.
15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47.
16. Zimmerman JM, Eliezer N, Simha R. The characterization of amino acid sequences in proteins by statistical methods. *J Theor Biol* 1968;**21**:170–201.
17. Grantham R. Amino acid difference formula to help explain protein evolution. *Science* 1974;**185**:862–4.
18. Bowman AW, Azzalini A. *R package 'sm': nonparametric smoothing methods*. University of Glasgow, UK, and Università di Padova, Italy; 2014.
19. World Health Organization. *Global tuberculosisreport 2015*. Geneva: WHO; 2015 p. 204.
20. Valentini D, Ferrara G, Advani R, Hallander HO, Maeurer MJ. Serum reactome induced by *Bordetella pertussis* infection and pertussis vaccines: qualitative differences in serum antibody recognition patterns revealed by peptide microarray analysis. *BMC Immunol* 2015;**16**:40.
21. Valentini D, Gaseitsiwe S, Maeurer M. Humoral 'reactome' profiles using peptide microarray chips. *Trends Immunol* 2010;**31**:399–400.
22. Rao M, Valentini D, Poiret T, Dodoo E, Parida S, Zumla A, et al. B in TB: B cells as mediators of clinically relevant immune responses in tuberculosis. *Clin Infect Dis* 2015;**61**(Suppl 3):S225–34.
23. Kang HE, Weng CC, Saijo E, Saylor V, Bian J, Kim S, et al. Characterization of conformation-dependent prion protein epitopes. *J Biol Chem* 2012;**287**:37219–32.
24. Savage PA. Tumor antigenicity revealed. *Trends Immunol* 2014;**35**:47–8.
25. Lustrek M, Lorenz P, Kreutzer M, Qian Z, Steinbeck F, Wu D, et al. Epitope predictions indicate the presence of two distinct types of epitope-antibody-

reactivities determined by epitope profiling of intravenous immunoglobulins. *PLoS One* 2013;**8**:e78605.
26. Deng W, Xie J. Ins and outs of *Mycobacterium tuberculosis* PPE family in pathogenesis and implications for novel measures against tuberculosis. *J Cell Biochem* 2012;**113**:1087–95.
27. Bhat KH, Das A, Srikantam A, Mukhopadhyay S. PPE2 protein of *Mycobacterium tuberculosis* may inhibit nitric oxide in activated macrophages. *Ann NY Acad Sci* 2013;**1283**:97–101.
28. Voskuil MI, Schnappinger D, Rutherford R, Liu Y, Schoolnik GK. Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. *Tuberculosis* 2004;**84**:256–62.
29. Manganelli R, Voskuil MI, Schoolnik GK, Smith I. The *Mycobacterium tuberculosis* ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. *Mol Microbiol* 2001;**41**:423–37.
30. Walters SB, Dubnau E, Kolesnikova I, Laval F, Daffe M, Smith I. The *Mycobacterium tuberculosis* PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. *Mol Microbiol* 2006;**60**:312–30.
31. van Soolingen D, de Haas PE, Hermans PW, Groenen PM, van Embden JD. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. *J Clin Microbiol* 1993;**31**:1987–95.
32. Solans L, Gonzalo-Asensio J, Sala C, Benjak A, Uplekar S, Rougemont J, et al. The PhoP-dependent ncRNA Mcr7 modulates the TAT secretion system in *Mycobacterium tuberculosis*. *PLoS Pathog* 2014;**10**:e1004183.
33. Abou-Zeid C, Garbe T, Lathigra R, Wiker HG, Harboe M, Rook GA, Young DB. Genetic and immunological analysis of *Mycobacterium tuberculosis* fibronectin-binding proteins. *Infect Immun* 1991;**59**:2712–8.
34. Espitia C, Laclette JP, Mondragon-Palomino M, Amador A, Campuzano J, Martens A, et al. The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins. *Microbiology* 1999;**145**(Pt 12):3487–95.
35. Sani M, Houben EN, Geurtsen J, Pierson J, de Punder K, van Zon M, et al. Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. *PLoS Pathog* 2010;**6**:e1000794.
36. 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.
37. Brodin P, Rosenkrands I, Andersen P, Cole ST, Brosch R. ESAT-6 proteins: protective antigens and virulence factors? *Trends Microbiol* 2004;**12**:500–8.
38. Volkman HE, Pozos TC, Zheng J, Davis JM, Rawls JF, Ramakrishnan L. Tuberculous granuloma induction via interaction of a bacterial secreted protein with host epithelium. *Science* 2010;**327**:466–9.
39. Pai M, Denkinger CM, Kik SV, Rangaka MX, Zwerling A, Oxlade O, et al. Gamma interferon release assays for detection of *Mycobacterium tuberculosis* infection. *Clin Microbiol Rev* 2014;**27**:3–20.
40. Kaufmann SH, Lange C, Rao M, Balaji KN, Lotze M, Schito M, et al. Progress in tuberculosis vaccine development and host-directed therapies—a state of the art review. *Lancet Respir Med* 2014;**2**:301–20.
41. Akif M, Khare G, Tyagi AK, Mande SC, Sardesai AA. Functional studies of multiple thioredoxins from *Mycobacterium tuberculosis*. *J Bacteriol* 2008;**190**:7087–95.
42. Zhang Z, Hillas PJ, Ortiz de Montellano PR. Reduction of peroxides and dinitrobenzenes by *Mycobacterium tuberculosis* thioredoxin and thioredoxin reductase. *Arch Biochem Biophys* 1999;**363**:19–26.
43. Wieles B, Ottenhoff TH, Steenwijk TM, Franken KL, de Vries RR, Langermans JA. Increased intracellular survival of Mycobacterium smegmatis containing the Mycobacterium leprae thioredoxin–thioredoxin reductase gene. *Infect Immun* 1997;**65**:2537–41.
44. Deshpande RG, Khan MB, Bhat DA, Navalkar RG. Isolation of a 43 kDa protein from *Mycobacterium tuberculosis* H37Rv and its identification as a pyridine nucleotide transhydrogenase. *J Appl Bacteriol* 1994;**77**:639–43.
45. Hutter B, Singh M. Properties of the 40 kDa antigen of *Mycobacterium tuberculosis*, a functional l-alanine dehydrogenase. *Biochem J* 1999;**343**(Pt 3):669–72.
46. Koch A, Mizrahi V, Warner DF. The impact of drug resistance on *Mycobacterium tuberculosis* physiology: what can we learn from rifampicin? *Emerg Microbes Infect* 2014;**3**(3):e17. doi:http://dx.doi.org/10.1038/emi.2014.17.
47. Manca C, Paul S, Barry 3rd CE3rd, Freedman VH, Kaplan G. *Mycobacterium tuberculosis* catalase and peroxidase activities and resistance to oxidative killing in human monocytes in vitro. *Infect Immun* 1999;**67**:74–9.
48. McCammon MT, Gillette JS, Thomas DP, Ramaswamy SV, Graviss EA, Kreiswirth BN, et al. Detection of rpoB mutations associated with rifampin resistance in *Mycobacterium tuberculosis* using denaturing gradient gel electrophoresis. *Antimicrob Agents Chemother* 2005;**49**:2200–9.