A standardized production pipeline for high profile targets from *Mycobacterium tuberculosis*

Morlin C. Milewski¹, Tobias Broger², Joanna Kirkpatrick³, Angela Filomena⁴, Dana Komadina¹, Nicole Schneiderhan-Marra⁴, Matthias Wilmanns¹,⁵ and Annabel H. A. Parret¹

¹ European Molecular Biology Laboratory (EMBL), Hamburg Unit, Hamburg, Germany
² Foundation for Innovative New Diagnostics (FIND), Geneva, Switzerland
³ European Molecular Biology Laboratory (EMBL), Proteomics Core Facility, Heidelberg, Germany
⁴ Natural and Medical Sciences Institute, University of Tübingen, Reutlingen, Germany
⁵ University of Hamburg Clinical Center Hamburg-Eppendorf, Hamburg, Germany

**Purpose:** Tuberculosis is still a major threat to global health. New tools and strategies to produce disease-related proteins are quintessential for the development of novel vaccines and diagnostic markers.

**Experimental design:** To obtain recombinant proteins from *Mycobacterium tuberculosis* (Mtb) for use in clinical applications, a standardized procedure was developed that includes subcloning, protein expression in *Mycobacterium smegmatis* and protein purification using chromatography. The potential for the different protein targets to serve as diagnostic markers for tuberculosis was established using multiplex immunoassays.

**Results:** Twelve soluble proteins from Mtb, including one protein complex, were purified to near-homogeneity following recombinant expression in *M. smegmatis*. Protein purity was assessed both by size exclusion chromatography and MS. Multiplex serological testing of the final protein preparations showed that all but one protein displayed a clear antibody response in serum samples from 278 tuberculosis patients.

**Conclusion and clinical relevance:** The established workflow comprises a simple, cost-effective, and scalable pipeline for production of soluble proteins from Mtb and can be used to prioritize immunogenic proteins suitable for use as diagnostic markers.

**Keywords:** Antigen / *Mycobacterium* / Multiplex immunoassay / Recombinant protein production

Additional supporting information may be found in the online version of this article at the publisher’s web-site

1 Introduction

Tuberculosis (TB) is far from being a disease of the past. Each year 1.5 million die as a result of TB, a highly infectious and complex illness caused by the airborne bacterium *Mycobacterium tuberculosis* (Mtb) [1]. The WHO End TB Strategy (http://www.who.int/tb/post2015_strategy/en/) aims to end the global TB epidemic, with targets to reduce TB deaths by 95% and incidence rate by 90% between 2015 and 2035. Moving forward to the 2035 targets requires the ensured availability of new tools, including (i) affordable and highly sensitive diagnostic tests for all forms of TB that can be implemented at the point of care, (ii) a vaccine that protects those of all ages who are not yet infected—preferably one that can also prevent people with latent TB from progressing to active disease and (iii) highly effective, shorter drug regimens, including regimens for TB infection caused by drug-resistant TB. The development of vaccines, diagnostics, and drugs depends upon a fundamental knowledge of biochemical pathways and intracellular processes critical for Mtb pathogenesis.

**Correspondence:** Dr. Annabel H. A. Parret, European Molecular Biology Laboratory, Hamburg Unit, Notkestraße 85, 22607 Hamburg, Germany

E-mail: parret@embl-hamburg.de

Fax: +49-40-89902149

**Abbreviations:** HIC, hydrophobic interaction chromatography; IEX, ion exchange chromatography; IgG, immunoglobulin G; Mtb, *Mycobacterium tuberculosis*; PE, protein extraction; TB, tuberculosis

[The copyright line of this article has been changed since first published on 3 August 2016 from the standard copyright to CC-BY].
Clinical Relevance

In 2014, tuberculosis (TB) killed 1.5 million people and now ranks alongside HIV as a leading cause of death worldwide. The causative agent of TB is the airborne bacterium Mycobacterium tuberculosis (Mtbt), which mainly attacks the lungs and is easily transmittable through inhalation of aerosol droplets. To open new opportunities for prevention and innovative therapies, it is essential to gain a better mechanistic understanding of the underlying molecular processes both within the pathogen and its interactions with the human host during infection. Progress toward understanding the structure and function of Mtbt proteins has, however, been hampered by a scarcity of easily applicable protocols and tools for producing sufficient amounts of recombinant protein for clinical applications, such as vaccine development and biomarker identification. Here, we describe an effective expression and standardized expression workflow to rapidly assess whether functional Mtbt proteins can be produced via recombinant expression in nonpathogenic Mycobacterium smegmatis. Our results show that M. smegmatis is a valuable host for the efficient production of immunogenic proteins from Mtbt, which may advance the development of tools for better diagnosis, prevention, and treatment of TB.

In 2014, tuberculosis (TB) killed 1.5 million people and now ranks alongside HIV as a leading cause of death worldwide. The causative agent of TB is the airborne bacterium Mycobacterium tuberculosis (Mtbt), which mainly attacks the lungs and is easily transmittable through inhalation of aerosol droplets. To open new opportunities for prevention and innovative therapies, it is essential to gain a better mechanistic understanding of the underlying molecular processes both within the pathogen and its interactions with the human host during infection. Progress toward understanding the structure and function of Mtbt proteins has, however, been hampered by a scarcity of easily applicable protocols and tools for producing sufficient amounts of recombinant protein for clinical applications, such as vaccine development and biomarker identification. Here, we describe an effective expression and standardized expression workflow to rapidly assess whether functional Mtbt proteins can be produced via recombinant expression in nonpathogenic Mycobacterium smegmatis. Our results show that M. smegmatis is a valuable host for the efficient production of immunogenic proteins from Mtbt, which may advance the development of tools for better diagnosis, prevention, and treatment of TB.

2. Materials and methods

2.1 Protein expression in M. smegmatis

Expression vectors, bacterial strains and growth conditions used in this study are described in the Supporting Information Table 1. M. smegmatis starter cultures were cultivated from freshly streaked colonies or glycerol stocks for 3 days at 37°C. For small-scale and large-scale expression studies, 1% of a starter culture was used to inoculate 100 mL or 1 L 7H9 expression medium, respectively. When cultures had reached an optical density of ~2.5 measured at 600 nm, protein expression was induced with 34 mM acetamide (Sigma-Aldrich, Germany) and cultivation was executed for an additional 16–24 h. Cells were harvested by centrifugation at 3220 × g or 7550 × g for small-scale and large-scale purification, respectively. Pellets were flash-frozen in liquid N2 and stored at −80°C until further use.

2.2 Protein purification

In brief, crude extracts were obtained by centrifugation for 45 min at 4°C (40 000 × g) after sonication of the cells in 20–25 mL protein extraction (PE) buffer (300 mM NaCl, 50 mM HEPES pH 8.0, 20 mM imidazole) as described previously [13]. The supernatant was filtered through a 0.45 μm filter and applied onto a Poly-Prep column (Bio-Rad, Germany) containing 200 μL (for 50 mL purification) or 1 mL (for 1 L purification) Ni-NTA agarose (Qiagen, Germany) preequilibrated with PE buffer. Protein contaminants were removed by multiple washing steps using PE buffer. His-tagged proteins were eluted with 3–5 column volumes PE buffer containing 500 mM imidazole. All purification fractions were collected and analyzed on NuPAGE Novex 4–12% Bis-Tris gels (Life Technologies GmbH, Germany). SDS-PAGE gels were stained with Instant Blue gel stain (Expedeon, Cambridge, UK) that has a sensitivity of 5–25 ng protein/band. Protein size was verified by comparison with the Rotimark 10–150 protein marker (Carl Roth, Germany). For final protein preparations, recombinant protein was dialyzed into SEC buffer (50 mM HEPES pH 8.0, 100 mM NaCl)
and passed through a SEC column (HiLoad 16/60 Superdex 75 or 200, GE Healthcare Life Sciences, Germany) using SEC buffer as running buffer. SEC columns were calibrated using a mixture of four proteins, thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), and myoglobin (17 kDa) (Gel Filtration Standards mixture, Bio-Rad, Richmond, CA, USA). Protein purity was verified by SDS-PAGE and fractions containing pure protein were pooled, flash-frozen in liquid N₂ and stored at −80°C. When additional chromatography steps were necessary in order to achieve 90% protein purity, protein samples were dialyzed or desalted using PD-10 columns (GE Healthcare Life Sciences) into ion exchange (IEX) buffer (50 mM HEPES pH 8.0, 100 mM NaCl) or hydrophobic interaction chromatography (HIC) starting buffer (50 mM HEPES pH 8.0, 1 M (NH₄)₂SO₄). For IEX, proteins were applied to a MonoQ 5/50 GL column (GE Healthcare Life Sciences) and eluted with a linear gradient of 10 column volumes from 100 mM to 500 mM NaCl. Fractions containing the target protein were identified by SDS-PAGE, pooled, and protein buffer was exchanged to SEC buffer or HIC buffer depending on the necessary subsequent step. Proteins to be further purified by HIC were applied to a 1 mL Phenyl Sepharose HP column (GE Healthcare Life Sciences) and eluted with a linear gradient of 100–0% (NH₄)₂SO₄. Additional details with respect to protein purification are described in the supplementary information. Protein sequences are given in Supporting Information Fig. 1.

### 2.3 Mass spectrometry

Intact protein molecular weight analysis was performed using LC-MS on a Q-TOF mass spectrometer (Waters GmbH, Germany). Protein samples in solution were acidified and subjected to C₄ reverse phase chromatography before electrospray ionization in positive ion mode. Raw data charge state series spectra were deconvoluted to neutral molecular weight using a maximum entropy algorithm. Protein digestion was carried out either in-gel or in-solution with different enzymes and the resulting peptides were analyzed by nano-LC-MS/MS on an Orbitrap mass spectrometer (Fisher Scientific, Germany). Data processing was performed using Mascot (Matrix Science Ltd., UK) as the search engine. Detailed protocols are provided as supplementary information. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [14] partner repository with the dataset identifier PXD004133.

### 2.4 Multiplex immunoassay

The functionality of the proteins was tested by measuring immunoglobulin G (IgG) antibody response in serum samples from 445 well-characterized TB suspects from Vietnam and Peru (167 Non-TB and 278 TB) as described previously [15]. Specimens were provided by the Foundation of Innovative New Diagnostics (FIND) and tested at the Natural and Medical Sciences Institute (NMI, University of Tübingen, Reutlingen, Germany). In brief, the purified proteins were covalently coupled to color-coded beads (MagPlex Microspheres, Luminex Corp, Austin, USA). The bead mixture was incubated with patient sera and protein-bound human antibodies were detected with PE-labeled anti-human IgG on a Luminex FlexMAP3D (Luminex Corp). Median fluorescence intensity was calculated for every sample based on >60 measured beads per bead sort and z-scores were used to illustrate antibody responses per patient in relationship to the mean of all patients. Receiver operating characteristic (ROC) curves were used to illustrate the discriminating utility of the best performing protein and Mann–Whitney U-test was used to determine whether the antibody reactivity in the two populations (TB and Non-TB) is significantly different.

### 3 Results and discussion

Our overall goal was to develop a standard purification protocol with a minimum number of steps facilitating the production of milligram-amounts of proteins of at least 90% purity for subsequent testing of antibody binding in serum samples from TB suspects. Purity guidelines were guided by typical sample requirements for downstream applications, such as structural analysis and serodiagnostic assays.

#### 3.1 Target selection, construct design, and cloning of *M. tuberculosis* targets proteins

A panel of 18 diverse Mtb proteins previously associated with antibody reactivity in sera from TB patients [16] were selected to validate our procedure (Table 1). Targets were distributed across four functional groups, namely lipid metabolism; cell wall and cell processes; intermediary metabolism and respiration or virulence, detoxification, adaptation. Most proteins were either associated with membrane preparations [17] and/or culture supernatant filtrate [18–21]. At least eight proteins (PstS1, LprG, Acr, EspC, EspA, Ag85a, EsxA, EsxB) represented known virulence factors of the Mtb complex (reviewed in [5]), several of which play an essential role in Mtb pathogenicity. Molecular weights of selected target proteins ranged between 7.7 and 39.9 kDa with a pI varying from 4.5 to 7.8. The heterodimeric complex of EsxB and EsxA (further referred to as EsxB/A) was included as an internal reference as it was shown previously to be efficiently produced in *M. smegmatis* [22].

For the well-established and most commonly used *E. coli* expression platform, an extensive catalog of molecular tools is available, which greatly simplify recombinant protein production [7]. In contrast, expression vectors and protocols geared toward high-yield expression in mycobacteria are rather limited (reviewed in [9]). In this study, we used in-house
Table 1. *M. tuberculosis* H37Rv proteins expressed in *M. smegmatis* groEL1ΔC

| Gene number | Protein namea | Uniprot entryb | Protein identity | Functional groupc | Subcellular locationd | Molecular mass (kDa)e | Isoelectric pointf |
|-------------|---------------|----------------|------------------|-------------------|-----------------------|----------------------|------------------|
| Rv0632c     | EchA3         | P96907         | Enoyl-CoA hydratase | 1                 | n.a.                  | 24.4                 | 5.52             |
| Rv0934      | PstS1         | P9WGU1         | Phosphate-specific transport substrate-binding protein-1 | 2                 | M; S[18]              | 38.2                 | 5.14             |
| Rv1411c     | LprG          | P9WK45         | Lipoarabinomannan carrier protein | 2                 | M; S[19]              | 24.5                 | 7.78             |
| Rv1837c     | GIClB         | P9WK17         | Malate synthase G | 3                 | C; S[18]              | 80.4                 | 5.03             |
| Rv1860      | Apa (Mpt32)   | P9WIR7         | Alanine and proline-rich secreted protein | 2                 | S[18]                 | 32.7                 | 4.93             |
| Rv1886c     | Ag85b (FbpB)  | P9WQP1         | Diacylglycerol acyltransferase/mycolyltransferase | 1                 | S[18]                 | 34.6                 | 5.62             |
| Rv1980c     | Mpt64         | P9WIN9         | Immunogenic protein | 2                 | S[18]                 | 24.9                 | 4.84             |
| Rv2031c     | Acr (HspX)    | P9WMK1         | Alpha-crystallin | 0                 | M; C                  | 16.2                 | 5.00             |
| Rv2654c     | Antitoxin Rv2654 | P9WJ11     | Antitoxin component of a toxin-antitoxin module (Rv2654c-Rv2653c) | 0g               | n.a.                  | 7.7                  | 5.04             |
| Rv2873      | Mpt83         | P9WNF3         | Immunogenic cell surface lipoprotein | 2                 | M; S[18]              | 22.1                 | 4.86             |
| Rv3615c     | EspC          | P9WJD7         | ESX-1 secretion-associated protein | 2                 | M; S                  | 10.8                 | 5.10             |
| Rv3616c     | EspA          | P9WJE1         | ESX-1 secretion-associated protein | 2                 | M; S                  | 39.9                 | 5.19             |
| Rv3804c     | Ag85a (FbpA)  | P9WQP3         | Diacylglycerol acyltransferase/mycolyltransferase | 1                 | C; M; S[18]           | 35.7                 | 6.08             |
| Rv3841      | BfrB          | P9WNE5         | Ferritin          | 3                 | C; M; S[18]           | 20.4                 | 4.73             |
| Rv3864      | EspE          | P9WJD3         | ESX-1 secretion-associated protein | 2                 | M                      | 42.1                 | 4.72             |
| Rv3874      | EsxB          | P9WNK5         | ESAT-6-like protein CFP-10 | 2                 | S[18]                 | 9.9                  | 4.59             |
| Rv3875      | EsxA          | P9WNK7         | 6 kDa early secretory antigenic target ESAT-6 | 2                 | S[18]                 | 10.8                 | 4.48             |
| Rv3881c     | EspB          | P9WJD9         | ESX-1 secretion-associated protein | 2                 | S[18]                 | 47.6                 | 4.75             |

a) Commonly used alternative names are indicated in brackets.
b) Protein accession number in UniProt database (http://www.uniprot.org/) version July 2016.
c) Explanation of functional group: (0) virulence, detoxification, adaptation, (1) lipid metabolism, (2) cell wall and cell processes, (3) intermediary metabolism and respiration. Functional group codes are taken from the web server (http://genolist.pasteur.fr/TubercuList/).
d) The subcellular location of each target as indicated in the UniProt database. For certain targets additional data on subcellular localization was retrieved from selected publications.
e) Molecular weights of apo-proteins without protein tags as reported in the Uniprot database.
f) Isoelectric points were calculated using the ProtParam tool (http://web.expasy.org/protparam/).
g) This protein is wrongly functionally categorized as “insertion sequences and phages.” According to Ramage et al., protein Rv2654 is the antitoxin component of a toxin-antitoxin module [34] and has therefore been categorized in this study as “virulence, detoxification, and adaptation protein.”

...generated inducible expression vectors, pMyNT and pMyC [22]. These *E. coli*/mycobacterial shuttle vectors contain an acetamidase promoter [23–25], a hygromycin resistance gene and an N- or C-terminal hexahistidine-tag in case of pMyNT or pMyC, respectively. Polyhistidine-tags substantially facilitate protein purification, while rarely adversely affecting protein structure or function [26]. Using IMAC, polyhistidine-tagged proteins can be enriched 100-fold in a single step resulting in high yields of up to 95% pure protein [27]. In case the affinity tag would influence protein solubility or function, pMyNT contains a recognition site for tobacco etch virus protease allowing for its efficient removal from the recombinant protein by proteolytic cleavage.

One to three constructs were designed per target gene including the full-length gene sequence. For targets reported to be secreted or associated with the Type VII secretion system we prepared additional constructs to express the apo-proteins with a C-terminal hexahistidine-tag. For six targets with experimentally validated N-terminal signal sequences [28], truncated constructs were created in order to produce mature protein.

### 3.2 Expression screening and solubility assessment

Following sequence verification, selected constructs were transformed to *M. smegmatis* groEL1ΔC by electroporation (Fig. 1). This modified expression strain was generated in our laboratory and is ideally suited for expression of polyhistidine-tagged proteins due to a deletion of the...
Figure 1. Schematic overview of established workflow used in this study to produce recombinant proteins from *M. tuberculosis* in *M. smegmatis*. Essential checkpoints throughout the workflow are indicated. (A) Workflow for small-scale expression and solubility assessment from 100 mL *M. smegmatis* cultures. (B) Scale-up of protein purification for promising Mtb targets including a polishing step using SEC and protein identification by MS analysis.

Histidine-rich C-terminal tail of the common contaminant GroEL1, thus consequently increasing the speed and efficiency of protein purification using IMAC [13]. Initial expression experiments were performed in 100 mL cultures in rich growth medium, where expression was induced for 24 h. Cells from 50 mL cultures were lysed in a standardized lysis buffer and soluble protein was recovered after high-speed centrifugation. The lysis step is the only step in our workflow where targets cannot be handled simultaneously. Soluble proteins were further mixed with Ni-NTA agarose beads to capture polyhistidine-tagged proteins in batch-binding mode using reusable plastic columns. The IMAC step was designed in such a way that the complete experiment could be carried out on a work bench without the need of specialized equipment. In this way, lysates from six constructs could be easily processed in parallel. Binding, washing, and elution of tagged proteins were achieved by gravity flow purification. Whole cell lysate and protein preparations were resolved by SDS-PAGE and the expression level, solubility, and IMAC recovery of target proteins was assessed by inspection of Coomassie-Blue-stained SDS-polyacrylamide gels. The levels of total expression for each target were scored visually using a standardized criterion; no or very low expression (no visible bands); medium (adequately sized band); high (very intense band). Expression of 24 out of 25 constructs was detectable by Coomassie staining (Fig. 2). Seventeen constructs, corresponding to 15 unique targets, yielded highly expressed protein, and nine constructs expressed moderately. Given
that our aim was to select promising constructs for high yield protein production from cellular extracts, low expressing constructs beyond our limits of detection were not considered further.

All proteins ran at the expected molecular weight on SDS-PAGE, with the exception of the proline-rich Apa protein (Supporting Information Fig. 3). For both constructs of Apa, an aberrant migratory behavior on SDS-PAGE was observed. However, proteins with high-proline content are notorious for running slower than their actual molecular weight during SDS-PAGE [29]. Alternatively, the presence of carbohydrate appendages could also result in a higher apparent molecular weight. However, this was ruled out by subsequent MS characterization (further details see below).

The solubility of each construct was further evaluated by comparing equal amounts of total protein before and after removal of insoluble material. Protein constructs were categorized as being either soluble (corresponding amounts of protein in total versus soluble protein fractions), insoluble (protein only detected in total protein fraction) or partly soluble (Fig. 2). Overall, constructs expressing well also yielded highly soluble protein. EspC and EspE were not pursued for further analysis due to low expression and solubility combined. Expression of another four full-length constructs (Ag85a<sub>opt</sub>, Ag85b, Mpt83<sub>opt</sub>, PstS1<sub>opt</sub>) was highly efficient, but did not yield soluble protein using standard buffer conditions for cell lysis. Further screening and optimization of cell lysis conditions would be necessary in order to rescue these particular targets. For our study, we rather opted to purify the corresponding proteins without a signal peptide.

Finally, the recovery after IMAC was evaluated by analyzing the proteins that eluted from the Ni-NTA resin at high imidazole concentrations. In this study, all targets that were shown to be soluble under the conditions evaluated, could also purified by IMAC, with the exceptions of Apa<sub>opt</sub>.

### 3.3 Large-scale purification of selected targets and quality control

On the basis of the expression levels, solubility, and purification criteria described above, 14 constructs were selected for upscaling, corresponding to 15 unique Mtb targets. Figure 2 summarizes the purification pipeline for each target as well as the obtained yield per liter of expression culture. For each target one construct was selected, exhibiting the most promising protein yield in large-scale purification. Three ESX-1–associated proteins (EspA, EspC, EspE) were not deemed suitable for upscaling as part of this investigation. In order to purify large quantities of these proteins, further optimization of expression and purification protocols would be indispensable.

To this end, recombinant proteins were produced in 1–2 L bacterial cultures and purified from contaminants by upscaling the methods employed in the small-scale screening using identical protein extraction and IMAC buffers (Fig. 1B). In order to achieve the defined sample requirements, in some cases an IEX or HIC step was included in order to remove protein contaminants (Supporting Information Table 2). Final polishing was achieved by SEC, which allows the removal of unwanted protein aggregates (Supporting Information Fig. 2). This technique is considered to be a crucial quality control method in the field of protein production [30, 31]. SEC analysis established conformational homogeneity of all produced proteins. Comparison of the SEC elution profiles with the elution profile of a set of known molecular mass standards, allows a rough estimation of the oligomeric state of the obtained proteins. Four proteins eluted at the expected retention volume corresponding to monomeric protein (EsxB, GlcB, Mpt64<sub>opt</sub>, PstS1<sub>opt</sub>), while both Ag85a<sub>1-244</sub> and Ag85b<sub>1-241</sub>...
migrated somewhat slower than expected in case of a globular protein of 34 and 33 kDa, respectively. This phenomenon could be explained by retention of the proteins due to non-specific binding to the SEC column. Six proteins (\text{Acr}_\text{opt}, \text{Ap}_\text{opt}, \text{BfrB}, \text{EchA3}, \text{EspB}_\text{opt}, \text{LprG}_\text{opt}) eluted from the SEC column earlier than globular proteins of corresponding molecular weights. This could be due to an extended protein structure resulting in a larger hydrodynamic radius and hence faster elution rate through the SEC column. Alternatively, it could indicate that the protein exists as stable oligomer in solution. The latter option clearly applies to BfrB. The derived molecular mass of BfrB (± 460 kDa) is consistent with the observation that this protein exists as a 24-mer in solution [32].

It should be noted that techniques such as dynamic light scattering or multiangle static light scattering combined with SEC would provide a far more accurate determination of the molecular weight of proteins. However, such an advanced biophysical analysis was beyond the scope of this study.

### 3.4 Mass spectrometry

To verify the identity and purity of the produced proteins, we performed intact mass measurements by LC-MS (Supporting Information Table 3). Intact protein masses matched the theoretical molecular weights for all proteins within the mass accuracy limits of the instrument (1–2 Da). No evidence for PTMs was found, e.g. phosphorylation or glycosylation. However, the intact mass of GlcB appeared 162 Da higher than its theoretical mass. A more in-depth MS analysis would be necessary to determine the underlying reason for this mass difference. Two proteins were not amenable to intact mass determination (EsxA, LprGopt), but protein identity was confirmed by follow-up peptide mass fingerprinting analysis (Supporting Information Table 4). However, in case of LprGopt, it should be noted that the MS analysis indicated that this protein preparation was quite significantly contaminated with endogenous \textit{M. smegmatis} proteins, as can also be seen in the final SDS-PAGE analysis (Supporting Information Fig. 3).

### 3.5 Antibody detection in serum samples

We investigated the immune response against proteins produced in our workflow in TB patients using a multiplex immunoassay. To compare assay performance to WHO targets (http://www.who.int/tb/publications/tpp_report/en/) sensitivity was calculated at a preset 95% specificity (targeting a TB detection test) and specificity was calculated at a preset sensitivity of 90% (targeting a TB triage test). A significantly higher antibody response in the TB compared to the non-TB group was shown for 11 out of 12 (92%) proteins, confirming binding of most proteins to human IgG (Table 2). Ag85b reached the highest sensitivity (36%) among all proteins at a specificity of 95%. Figure 3 shows antibody profiles against all proteins in this study (most significant proteins on top) for all 445 patients to visually

| Protein | Mann–Whitney U test (p-value) | ROC-AUC$^{a,b}$ | Detection test sensitivity (at 95% specificity) | Triage test specificity (at 90% sensitivity) |
|---------|-------------------------------|-----------------|-----------------------------------------------|--------------------------------------------|
| WHO target performance | c | > 0.90 | > 0.65 | > 0.70 |
| Ac$_{opt}$ | $6.15 \times 10^{-10}$ | 0.68 | 0.26 | 0.17 |
| EsxA | $1.88 \times 10^{-9}$ | 0.67 | 0.20 | 0.22 |
| LprG$_{opt}$ | $5.20 \times 10^{-9}$ | 0.67 | 0.18 | 0.24 |
| Ag85b$_{341}$ | $6.81 \times 10^{-9}$ | 0.66 | 0.36 | 0.18 |
| BfrB | $7.37 \times 10^{-9}$ | 0.66 | 0.16 | 0.21 |
| Ag85a$_{344}$ | $1.82 \times 10^{-8}$ | 0.66 | 0.32 | 0.16 |
| Mpt64$_{opt}$ | $1.81 \times 7$ | 0.65 | 0.34 | 0.17 |
| A$_{opt}$ | $4.26 \times 7$ | 0.64 | 0.23 | 0.20 |
| PstS1$_{24}$ | $5.35 \times 7$ | 0.64 | 0.12 | 0.21 |
| EchA3 | $4.32 \times 5$ | 0.62 | 0.17 | 0.14 |
| EspB$_{opt}$ | $4.58 \times 5$ | 0.62 | 0.12 | 0.16 |
| GlcB | $3.24 \times 10^{-d}$ | 0.53 | 0.10 | 0.14 |

- a) The diagnostic performance of two hypothetical TB tests (detection test and triage test) as they are targeted by the WHO (http://www.who.int/tb/publications/tpp_report/en/) were calculated as follows: sensitivity values of a detection test having a high specificity of 95% and specificity of a triage test having a high diagnostic sensitivity of 90%, respectively.
- b) Area under the receiver operating characteristic curve.
- c) Not applicable.
- d) Not significant at a significance level of > 0.05.
illustrate the significantly higher antibody response in the TB group for the 11 discriminatory proteins. Acr\textsubscript{opt} displayed the most elevated antibody response. For GlcB, the IgG response was not significantly different in the TB versus the non-TB group, suggesting that either the purification did not result in a functional product or that there is no significant antibody response toward GlcB in the course of TB infection. Additional functional and biochemical analyses would be required to investigate this in more detail.

4 Concluding remarks

We presented here an efficient workflow for screening Mtb targets for production as recombinant proteins in the related fast-growing expression host \textit{M. smegmatis}. As the proteins were of particular functional interest, considerable effort was made to ensure targets moved through the production pipeline with the aim to obtain as many targets as possible with at least 90% purity. Twelve targets were purified successfully from 1–2 L bacterial cultures following a standardized workflow where possible. Using this approach, we were able to rapidly select the optimal expression constructs for selected Mtb targets. All proteins were delivered to the Foundation for Innovative New Diagnostics (FIND) to assess their suitability as diagnostic markers for TB. Initial data showed increased antibody binding in the TB versus the non-TB group, but relatively low diagnostic performance. In follow-up experiments a comprehensive analysis of host-dependent protein modifications versus immunological differences (e.g. a direct comparison with native proteins purified from Mtb or \textit{E. coli}) will shed light on this issue. In this respect, it would be very valuable to extend the biophysical and biochemical characterization of the most promising protein targets in order to identify possible cooccurring protein species that could explain the relatively low diagnostic performance [33]. Furthermore, the performance and the feasibility of measuring IgG antibody response in combination with other markers await further efforts. Finally, our study provides a set of standardized protocols for rapid screening of Mtb proteins amenable for structural analysis.

We are indebted to Rosemary Wilson for the generation of Fig. 1. The authors acknowledge the EMBL’s Proteomics Core facility for the mass spectrometry analysis and Kate Beckham for reading the manuscript and providing critical feedback. This work was supported by funds from the Foundation for Innovative New Diagnostics (FIND).

The authors have declared no conflict of interest.

5 References

[1] Organization, W. H., Global Tuberculosis Report 2014. \textit{World Health Organization} 2014.
[2] Lienhardt, C., Fundamental research is the key to eliminating Tb. \textit{Nature} 2014, 507, 401.
[3] Schubert, O. T., Aebersold, R., Microbial proteome profiling and systems biology: applications to \textit{Mycobacterium tuberculosis}. \textit{Adv. Exp. Med. Biol.} 2015, 883, 235–254.
[4] Bell, C., Smith, G. T., Sweredoski, M. J., Hess, S., Characterization of the \textit{Mycobacterium tuberculosis} proteome by liquid chromatography mass spectrometry-based proteomics techniques: a comprehensive resource for tuberculosis research. \textit{J. Proteome Res.} 2012, 11, 119–130.
[5] Forrellad, M. A., Klepp, L. I., Gioffre, A., Sabio Y., Garcia, J. et al., Virulence factors of the \textit{Mycobacterium tuberculosis} complex. \textit{Virulence} 2013, 4, 3–68.
[6] Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M. et al., Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. \textit{Microbial. Cell Fact.} 2008, 7, 11.
Mycobacterium tuberculosis

Infect Immun. 1993, 11, 5

for structural studies is facilitated by the use of maltose
Elec-

Proc. Natl. Acad. Sci. USA 2000, Mycobacterium tubercu-

J. Anal. Methods

in the resistance to toxic compounds.
Gel Electrophoresis Of Proteins—A Practical
Plos One 2013, 8, E81753.

Mycobacterium smegmatis

[15] Planatscher, H., Rimmele, S., Michel, G., Potz, O. et al., Heterologous expression of mycobacterial esx complexes in Escherichia coli for structural studies is facilitated by the use of maltose binding protein fusions. Plos One 2013, 8, E81753.

[11] Triccas, J. A., Roche, P. W., Winter, N., Feng, C. G. et al., A 35-kilodalton protein is a major target of the human immune response to Mycobacterium leprae. Infect Immun. 1996, 64, 5171–5177.

[12] Garbe, T., Harris, D., Vordermeier, M., Lathigra, R. et al., Expression of the Mycobacterium tuberculosis 19-kilodalton antigen in Mycobacterium smegmatis: immunological analysis and evidence of glycosylation. Infect Immun. 1993, 61, 260–267.

[13] Noens, E. E., Williams, C., Anandhakrishnan, M., Poulsen, C. et al., Improved mycobacterial protein production using a Mycobacterium smegmatis groel1deltac expression strain. BMC Biotechnol. 2011, 11, 27.

[14] Vizcaíno, J. A., Csordas, A., Del-Toro, N., Dianes, J. A. et al., 2016 update of the pride database and its related tools. Nucleic Acids Res. 2016, 44, D447–D456.

[15] Planatscher, H., Rimmele, S., Michel, G., Potz, O. et al., Systematic reference sample generation for multiplexed serological assays. Sci. Rep. 2013, 3, 3259.

[16] Kunath-Velayudhan, S., Salamon, H., Wang, H. Y., Davidow, A. L. et al., Dynamic antibody responses to the Mycobacterium tuberculosis proteome. Proc. Natl. Acad. Sci. USA 2010, 107, 14703–14708.

[17] Gu, S., Chen, J., Dobos, K. M., Bradbury, E. M. et al., Comprehensive proteomic profiling of the membrane constituents of a Mycobacterium tuberculosis strain. Mol. Cell. Proteomics 2003, 2, 1284–1296.

[18] Malen, H., Berven, F. S., Fladmark, K. E., Wiker, H. G., Comprehensive analysis of exported proteins from Mycobacterium tuberculosis H37rv. Proteomics 2007, 7, 1702–1718.

[19] Bianco, M. V., Blanco, F. C., Imperiale, B., Forrellad, M. A. et al., Role of P27–P55 operon from Mycobacterium tuberculosis in the resistance to toxic compounds. BMC Infect Dis. 2011, 11, 195.

[20] Rosenkranz, I., King, A., Weldingh, K., Moniatte, M. et al., Towards the proteome of Mycobacterium tuberculosis. Electrophoresis 2000, 21, 3740–3756.

[21] Mattow, J., Schäuble, U. E., Schmidt, F., Hagens, K. et al., Comparative proteome analysis of culture supernatant pro-
teins from virulent Mycobacterium tuberculosis H37rv and attenuated M. Bovis Bcg Copenhagen. Electrophoresis 2003, 24, 3405–3420.

[22] Poulsen, C., Holton, S., Geerlof, A., Wilmanns, M., Song, Y. H., Stoichiometric protein complex formation and over-expression using the prokaryotic native operon structure. Fems Lett. 2010, 584, 669–674.

[23] Triccas, J. A., Parish, T., Britton, W. J., Gicquel, B., An inducible expression system permitting the efficient purification of a recombinant antigen from Mycobacterium smegmatis. Fems Microbiol. Lett. 1998, 167, 6.

[24] Mahenthiralingam, E., Draper, P., Davis, E. O., Colston, M. J., Cloning and sequencing of the gene which encodes the highly inducible acetamidase of Mycobacterium smegmatis. J. Gen. Microbiol. 1993, 139, 575–583.

[25] Daugelat, S., Kowall, J., Mattow, J., Bumann, D. et al., The Rd1 proteins of Mycobacterium tuberculosis: expression in Mycobacterium smegmatis and biochemical characterization. Microbes Infection 2003, 5, 1082–1095.

[26] Zhao, X., Li, G., Liang, S., Several affinity tags commonly used in chromatographic purification. J. Anal. Methods Chem. 2013, 2013, 581093.

[27] Bornhorst, J. A., Falke, J. J., Purification of proteins using polyhistidine affinity tags. Methods Enzymol. 2000, 326, 245–254.

[28] De Souza, G. A., Leversen, N. A., Malen, H., Wiker, H. G., Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. J. Proteomics 2011, 75, 502–510.

[29] Hames, B. D., Gel Electrophoresis Of Proteins—A Practical Approach, 3rd edn., Oxford University Press, New York 1998.

[30] Raynal, B., Lenormand, P., Baron, B., Hoos, S., England, P., Quality assessment and optimization of purified protein samples: why and how? Microbial Cell Factories 2014, 13, 986.

[31] Structural Genomics Consortium, China Structural Genomics Consortium, Northeast Structural Genomics Consortium, Graslund, S. et al., Protein production and purification. Nat. Methods 2008, 5, 135–146.

[32] McMeth, L. M., Habel, J. E., Sankaran, B., Yu, M. et al., Crystalization and preliminary X-ray crystallographic analysis of a Mycobacterium tuberculosis ferritin homolog, Bfrb. Acta Crystallographica. Sect. F Struct. Biol. Cryst. Commun. 2010, 66, 1657–1661.

[33] Jungblut, P. R., Thiede, B., Schluter, H., Towards deciphering proteomes via the proteoform, protein speciation, moonlighting and protein code concepts. J. Proteomics 2016, 134, 1–4.

[34] Ramage, H. R., Connolly, L. E., Cox, J. S., Comprehensive functional analysis of Mycobacterium tuberculosis toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. Plos Genet. 2009, 5, E1000767.