SIGNIFICANCE

This is the first study that compares the saliva and the sputum from active TB patients and their contacts. Our findings strongly suggest that TB patients show not only an activation of processes that are related to complement activation and modulation of inflammation, but also an imbalance in carbohydrate and lipid metabolism. In addition, those individuals who do not get infected after direct exposure to the pathogen display a typical proteomic signature in the sputum which is a reflection of the secretion from the nasal and oral mucosa, the first immunological barriers that *M. tuberculosis* encounters in the host. Thus, this result indicates the importance of the processes related to the innate immune response in fighting the initial events of the infection.
Mycobacterium tuberculosis

Shotgun Quantitative Proteomics

Specific Protein Signature of

Tuberculosis patients:
High levels of markers of inflammation and acute-phase response vs. latent-infected contacts in both fluids

Uninfected contacts:
Increased levels of proteins related to innate immune response vs. latent-infected contacts in sputum

TUBERCULOSIS PATIENTS AND CONTACTS SAMPLE COLLECTION

-SPUTUM

-SALIVA

Mycobacterium tuberculosis

Graphical Abstract
HIGHLIGHTS

- Proteomic analysis of saliva and sputum in tuberculosis patients and contacts.
- Both fluids differ in terms of protein composition.
- Active TB patients show markers of inflammation and complement activation.
- Active TB patients present a decrease in enzymes related to sugar metabolism.
- Markers of innate immune response are higher in the sputum of uninfected contacts.
High-Resolution Quantitative Proteomics applied to the study of the specific protein signature in the sputum and saliva of active tuberculosis patients and their infected and uninfected contacts

Jesús Mateos¹*, Olivia Estévez², África González-Fernández², Luis Anibarro²,³,⁴, Ángeles Pallarés³, Rajko Reljic⁵, José M. Gallardo¹, Isabel Medina¹ and Mónica Carrera¹*

¹Spanish National Research Council (CSIC), Vigo, Pontevedra, Spain
²Biomedical Research Centre (CINBIO), Galician Singular Center of Research, Galicia
³Health Research Institute (IIS-GS), University of Vigo, Vigo, Pontevedra, Spain.
⁴Tuberculosis Unit, Infectious Diseases, Internal Medicine service, Complexo Hospitalario Universitario de Pontevedra, Galicia Sur Health Research Institute (IIS-GS, Pontevedra, Spain.
⁵Mycobacterial Infections Study Group (GEIM) of the Spanish Society of Infectious Diseases and Clinical Microbiology (SEIMC), Madrid, Spain.
⁶St. George’s, University of London, UK.

*CORRESPONDING AUTHOR: Jesús Mateos, PhD
Spanish National Research Council (CSIC), 36208, Vigo, Pontevedra, Spain.
Phone: +34 986231930
Fax: +34 986 292762
e-mail: jmateos@iim.csic.es

^CO-CORRESPONDING AUTHOR: Mónica Carrera, PhD
Spanish National Research Council (CSIC), 36208, Vigo, Pontevedra, Spain.
Phone: +34 986231930
Our goal was to establish panels of protein biomarkers that are characteristic of patients with microbiologically confirmed pulmonary tuberculosis (TB) and their contacts, including latent TB-infected (LTBI) and uninfected patients. Since the first pathogen-host contact occurs in the oral and nasal passages the saliva and sputum were chosen as the biological fluids to be studied. Quantitative shotgun proteomics was performed using a LTQ-Orbitrap-Elite platform. For active TB patients, both fluids exhibited a specific accumulation of proteins that were related to complement activation, inflammation and modulation of immune response. In the saliva of TB patients, a decrease of in proteins related to glucose and lipid metabolism was detected. In contrast, the sputum of uninfected contacts presented a specific proteomic signature that was composed of proteins involved in the perception of bitter taste, defense against pathogens and innate immune response, suggesting that those are key events during the initial entry of the pathogen in the host.
SIGNIFICANCE

This is the first study to compare the saliva and sputum from active TB patients and their contacts. Our findings strongly suggest that TB patients show not only an activation of processes that are related to complement activation and modulation of inflammation but also an imbalance in carbohydrate and lipid metabolism. In addition, those individuals who do not get infected after direct exposure to the pathogen display a typical proteomic signature in the sputum, which is a reflection of the secretion from the nasal and oral mucosa, the first immunological barriers that *M. tuberculosis* encounters in the host. Thus, this result indicates the importance of the processes related to the innate immune response in fighting the initial events of the infection.

HIGHLIGHTS

- Proteomic analysis of saliva and sputum in tuberculosis patients and contacts.
- Both fluids differ in terms of protein composition.
- Active TB patients show markers of inflammation and complement activation.
- Active TB patients present a decrease in enzymes related to sugar metabolism.
- Markers of innate immune response are higher in the sputum of uninfected contacts.
1. Introduction

Tuberculosis (TB) is a major health problem globally (1). TB is caused by the bacterium *Mycobacterium tuberculosis* (MTB). Roughly one third of the world's population carries MTB in a dormant form (2). TB is responsible for the death of more than 1.8 million people each year, making it one of the leading causes of mortality and the most common cause of death by a single infectious agent (3). Since evidence of the infection has been found in human remains dating back 5000 years, TB is considered one of the oldest diseases in mankind (4). Pulmonary TB is air transmitted from an active TB patient to a healthy individual through coughing or sneezing. Infection occurs through nasal/oral inhalation of aerosol droplets carrying MTB (5). The smaller droplets are able to reach the lower lung, and, after recruitment of macrophages and dendritic cells, they form the granuloma, which is a host-defensive structure that is characteristic of latent infections (6). Granulomas provide a fibrotic physical barrier between the infected, necrotic area and the healthy neighboring tissue (7). LTBI patients cannot infect a healthy individual; however, eventually, activation of the pathogen occurs in approximately 10% of the cases due to recurrent infections, immunosuppression or a weakened health state of the host (8).

Remarkably, many individuals in close contact with an active TB patient do not become infected. Increasing evidence suggests that genetic heterogeneity of the host affects immune response to intracellular pathogens, such as mycobacteria (9-11). Innate immune response is the first line of defense and plays a key role in the quality, strength and efficacy of the adaptive immune response (12). However, it remains unknown whether or not those healthy household contacts present a specific proteomic signature that could be linked to this particular state of special innate protection. Vaccines offer a
safe and cost-effective method to protect large populations against infectious diseases
or, alternatively, mitigate their clinical course (13). However, many poverty-related and
neglected infectious diseases such as TB continue to escape attempts to develop
effective vaccines against them (14). BCG (Bacille Calmette-Guerin), the vaccine
currently in use against TB, was developed eighty years ago and is widely used for
prevention, with an efficiency of more than 80% in children under 4 years (15).
Unfortunately, BCG efficiency in adolescents and adults is variable, ranging from 0 to
80% (16). The World Health Organization Global Strategy for the period 2015-2035
draws special attention to the urgent need to develop a new efficient vaccine against TB
(3, 17). Eliciting Mucosal Immunity in Tuberculosis (EMI-TB) consortium (www.emi-
tb.org) is a Horizon-2020-founded action that focuses on selecting candidates for
developing a new, nasal-administered vaccine against TB. Our laboratory (CSIC) is
integrated in the working-package 3, and our task is the proteomic analysis of biological
samples.

Proteomics provides a unique tool to analyze cellular and organism activity at
the protein level. Thus, proteomic profiling allows the elucidation of the links between
broad cellular pathways and individual molecules that were previously impossible to
predict using only traditional biochemical analysis. Biological fluids from human
subjects are a promising source for analyzing biological markers of health and disease
(18). Biological fluids contain biomolecules (including lipids, peptides, amino acids,
cytokines, proteases, enzymes, and antibodies) that present different physicochemical
properties. Since analytical proteomics has experienced extensive progress in the last
decade because of the emergence of mass-spectrometry-based techniques (MALDI-
TOF/TOF, LC-MS/MS), its combination with classical techniques for protein separation
(e.g., iso-electric focusing, chromatography, and two-dimensional gel electrophoresis)
facilitate the identification and characterization of thousands of proteins in a single
experiment. Proteomics is expected to be the tool of choice for the search for diagnostic or therapeutic biomarkers and for the identification and characterization of the proteins encoded by the genome (19, 20). One of the main objectives of proteomics is the identification of markers of disease by comparing the protein status in normal and pathological conditions.

Our specific aim for this study, as part of the Horizon2020 EMI-TB project, was to test two different biological fluids, sputum and saliva to establish the proteomic signature that is specific to active TB patients and their contacts, including LTBI and uninfected patients. We aim to determine indicators of the innate immune response that prevents the development of the disease and improve our understanding of the processes that are activated during MTB latency and active TB disease.
2. Material and methods

2.1. Patient selection and database management

Active culture-positive pulmonary TB patients and their close contacts were prospectively enrolled in patients attending the TB Unit in Pontevedra, Galicia, Spain (Table 1). Collection of samples initiated on September 2015, within the framework of the EMI-TB project (Project ID: 643558; Eliciting Mucosal Immunity to Tuberculosis; Ongoing project H2020-EU.3.1: SOCIETAL CHALLENGES; “Health, demographic change and well-being”). Exclusion criteria were ages less than 18 y, coinfection with the human immunodeficiency virus (HIV), any other immunosuppressive medical condition or concomitant use of immunosuppressive drugs. Patients with previous TB infection or LTBI were also excluded for the study. All patients accepted to be included in the study and signed informed consent. The database was elaborated using codes, keeping the anonymity of the person and including all the clinical fields considered relevant for the study. The informed consent was prepared, and all patients and contacts received a detailed explanation of the project and confidentiality. The total number of patients included in the study was 99, including 26 active TB patients and 73 contacts (Table 1). The contacts were screened using the tuberculin skin test (TST) and/or the Quantiferon-TB-Gold test (QFT) and again after 8-12 weeks after the last possible exposure to the index case if the first test was negative. Chest radiography was performed to exclude active TB in patients with a positive TST-QFT result. The patients were classified as follows: active TB, LTBI or uninfected, following Spanish national guidelines (21). A “Contact Score” was assigned to the contacts following this criteria: acid-fast bacilli (AFB) microscopy index: from 0 to 4; cavitary X-rays: No=0, Yes=1; household contact: No=0, Yes=1; hours of exposure/day: 0-3 hours: 1, 4-7 hours: 2, 8-
11 hours: 3, ≥ 12 hours: 4; type of exposure: Outdoors= 0.25, Different room: 1, As a bar: 2, As a class: 2.5, As an office: 3, As a room or car: 4; weeks in contact with Index Case: < 12 weeks: 0, ≥ 12 weeks: 1; sleeps in the same room: No=0, Yes=1; first-degree family relationship: No=0, Yes=1.

2.2. Collection and M. tuberculosis decontamination of saliva and sputum samples

Samples of saliva and sputum were collected from the selected patients in the Tuberculosis Unit of the “Complexo Hospitalario Universitario de Pontevedra” Hospital Facility following the ethical committee authorization from the Galician Ethics Committee for Clinical Research (CEIC, Ref 2014/492). Saliva was constituted only for a transparent liquid produced in the mouth without clots and was collected and processed following standard procedures in the hospital facilities (22). Not induced sputum samples (2-4 mL) were collected in sterile 50 mL polypropylene tubes and stored at 4°C until processing. After addition of four volumes of 1% β-mercaptoethanol in Phosphate Buffered Saline (PBS), the samples were shaken at 37°C for 15 min and centrifuged at 300 rpm for another 15 min to collect the supernatant fraction.

After collection, saliva and sputum samples were transferred to the Microbiology Unit of the same hospital, where they were processed for the inactivation of *Mycobacterium tuberculosis*. The mechanical disruption method was used to ensure the liquefaction of samples without damaging the proteins and RNA. For this purpose, zirconium-silica beads were used to homogenize the samples in a BeadBeater (Biospec, Bartlesville, OK 74005, USA). Once decontaminated, aliquots of the samples were seeded and cultured. The elimination of MTB from the treated samples was confirmed via microbial culture on Coletsos culture medium. This assay was performed in the Laboratory of Microbiology of the “Complexo Hospitalario de Pontevedra”. After
confirmation of the total elimination of the pathogen, frozen samples were transferred to the University of Vigo and stored at -80°C.

2.3. **Protein preparation**

Total protein from individual samples was determined by measuring the absorbance at 280 nm using the NanoDrop™ instrument (Thermo Fisher Scientific, San Jose, CA, USA) and was precipitated by adding six volumes of cold acetone and overnight incubation at -20°C. After centrifugation, the dried protein pellet was resuspended in 0.1 M triethylammonium bicarbonate (TEAB) buffer solution. The protein concentration for TMT labeling was determined using the Bicinchoninic Acid (BCA) assay (Sigma-Aldrich, St. Louis, MO, USA). Next, the protein integrity of all individual samples was confirmed via 10% acrylamide SDS-PAGE of 1 µg aliquots and silver staining of the resulting gels. The samples in which protein degradation was detected were discarded. The densitometry of the entire lane was used, if necessary, to correct protein quantification.

2.4. **Shotgun proteomics study design**

A schematic of the study design is presented in Figure 1A. Samples were selected for the quantitative proteomics study according to their protein concentration and integrity. Three complete Tandem Mass Tag (TMT) 10plex labeling procedures (Thermo Fisher Scientific) were performed for each biological fluid using individual samples. Within each TMT experiment, three active TB (channels 126, 127N and 127C), three infected LTBI (channels 128N, 128C and 129N) and three uninfected patients (channels 129C, 130N and 130C) were included, plus a standard sample resulting from mixing equal amounts of proteins for the nine samples included (channel 131). A summary of the samples used for the study is presented in Table 2.
2.5. **TMT 10plex labeling**

For the labeling, 100 µg of each individual sample was resuspended in a final volume of 100 µL of 0.1 M TEAB buffer solution, reduced/alkylated and digested with trypsin for 16 h at 37°C. Labeling with TMT 10plex reagents (Thermo Fisher Scientific) was performed following the manufacturer’s instructions. Briefly, LC-grade acetonitrile was used to dissolve the reagents (41 µL per 0.8 mg of reagent). After labeling for 1 hour and quenching with 8 µL of 5% hydroxylamine for 15 min, both steps at room temperature, all the channels were mixed in a single tube, aliquoted and dried in a speed-vac.

2.6. **Peptide fractionation by High-pH Reversed Phase**

Aliquots of 100 µg of the total labeled protein were reconstituted in 300 µL of trifluoroacetic acid (TFA), and the peptides were fractionated using the High-pH Reversed Phase fractionation kit (Thermo Fisher Scientific) following the manufacturer’s instructions. The peptide concentration in the resulting fractions (10 plus the Washing and the Flow-through) was determined via colorimetric analysis using the Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific).

2.7. **LC-MS/MS analysis and Orbitrap-Elite settings**

One microgram of the samples of each fraction, as determined using the colorimetric assay, were injected and analyzed in the Proteomics Facility of the University of Vigo (CACTI, Vigo, Spain) via LC-MS/MS using a Proxeon EASY-nLC II liquid chromatography system (Thermo Fisher Scientific) coupled to a LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). The peptides were separated on an RP column (EASY-Spray column, 50 cm × 75 µm ID, PepMap C18, 2 µm particles,
100 Å pore size, Thermo Fisher Scientific) with a 10 mm precolumn (Accucore XL C18, Thermo Fisher Scientific) using 0.1% formic acid (mobile phase A) and 98% ACN with 0.1% formic acid (mobile phase B). A 240 min linear gradient from 5% to 35% B was applied at a flow-rate of 300 nL per min. Ionization was performed in a nanosource using a spray voltage of 1.95 kV and a capillary temperature of 275°C. The peptides were analyzed in positive mode (1 μscan; 400–1600 amu), followed by 10 data-dependent HCD MS/MS scans (1μscans), using a normalized collision energy of 38% and an isolation width of 1.5 amu. Dynamic exclusion was enabled with a repeat count of 1, a repeat duration of 30 s, a duration of the exclusion of 80 s, and a relative exclusion width of 10 ppm. Unassigned charged ions were excluded from the analysis.

2.8. **Mass spectrometry data processing**

Raw data were loaded in the Xcalibur software (Thermo Fisher Scientific) for inspection of the chromatography profile and confirmation of the labeling of the peptides. Next, protein identification and quantification was performed using the Proteome Discoverer 2.1 software (Thermo Fisher Scientific). Peak lists were generated with a precursor signal-to-noise ratio of 1.5, and default settings were used to search the latest UniProtKB Release using the SEQUEST algorithm. The enzyme specificity was set to trypsin and one missed cleavage was tolerated. TMT-labeling and carbamidomethylation of cysteine were set as fixed modifications, whereas oxidation of methionine and N-terminal acetylation were set as variable modifications. The precursor ion mass tolerance was set to 7 ppm, and the product ion mass tolerance was set to 0.8 Da. A decoy database search was performed to determine the peptide false discovery rate (FDR) with the Target Decoy PSM Validator module. Quantification was performed using the Quantification Module, and normalization was performed against total peptide amount. A 1% peptide FDR threshold was applied.
Samples were categorized by the patient type (active TB, LTBI, uninfected and Internal Standard). Quantification jobs were alternatively launched using a) the Patient Type option for the global analysis and b) the individual ratio option for the nonparametric statistical analysis.

2.9. Selection of quantified proteins

Several filters were sequentially applied to the global quantification results to obtain the final list of quantified proteins: A) proteins quantified with at least two unique peptides, B) proteins quantified in the three TMTs for each biological fluid, C) p-value ≤ 0.0001 after performing the Kruskal-Wallis statistical test for all the different ratios. A summary of the filtering process is presented in Figure 1B.

2.10. Statistical analysis of the modulated proteins

The normalized ratios for the all the proteins quantified in the three TMTs with at least two unique peptides were extracted from the Proteome Discoverer 2.1 software using the “export to Excel” option and were used for the Kruskal-Wallis statistical analysis using the R software. Briefly, the normalized protein ratios (27 for each comparison) were imported into R commander console and represented in box-diagrams. For each protein, data were used to analyze the differences between the three studied ratios (27 ratios uninfected/LTBI, 27 ratios uninfected/TB and 27 ratios LTBI/TB) by applying a Kruskal-Wallis test. Differences in the modulation were considered as significant when p-value ≤ 0.0001.

The final list (UniProtKB accession number) of modulated proteins was analyzed using the String 10.1 software for the determination of pathways and biological processes modulated (p-value ≤ 0.001) in each group of patients.
3. Results

3.1. Sputum and saliva present differences in terms of protein composition and quantification

Proteomic datasets are deposited at the MassIVE repository (www.massive.ucsd.edu). Raw and processed files (EMI_TB_PROTEOMICS_CSIC; #MSV000081574) are public and freely accessible. A total of 1218 and 847 proteins were identified with at least one unique peptide in the sputum and saliva samples, respectively. Among them, 766 and 562 were quantified with at least two unique peptides. A total of 755 proteins were identified in both biological fluids (Supplementary Figure 1A). The sputum appeared to be more complex than the saliva, presenting a high number of specific protein isoforms, 548, whereas only 164 proteins were detected exclusively in the saliva. Additionally, the sputum and saliva appeared to differ in protein composition, because we found no correlation between similar TMT ratios when comparing both fluids (Supplementary Figure 1B, C and D).

3.2. Specific protein signature of active TB patients and contacts in sputum

For statistical analysis only those proteins were selected that were quantified in the three TMTs with at least two unique peptides (Figure 1B). A nonparametric Kruskal-Wallis test (supplementary information) was applied to detect statistically significant differences in the individual ratios (27 uninfected/LTBI, 27 TB/LTBI and 27 uninfected/TB). The results are represented as volcano-plots (Figure 2 and 6). Those proteins presenting a minimal 1.5-fold change and a p-value ≤ 0.0001 were considered as modulated.
Thus, we decided to investigate if the sputum presents a characteristic proteomic signature for each group of patients. The volcano plots suggested that there is an accumulation of a small subset of proteins in active TB versus LTBI patients (Figure 2A, right). Examples of these proteins are haptoglobin, alpha-1-acid glycoproteins 1 and 2 and fibrinogen. The list of significant proteins that are abundant in active TB patients was analyzed using the String 10.1 software, which indicated a strong interaction network (Figure 3A). Statistical pathway analysis showed (p-value ≤ 0.001) that most of the proteins are related to platelet degranulation, inflammation and acute phase response (Supplementary information 2). On the other hand, String 10.1 analysis of the proteins that were decreased in TB patients versus LTBI patients (Figure 2A, left) showed a weak interaction network (Figure 3B) between proteins related to endopeptidase activity and taste perception.

Uninfected patients exhibited an accumulation of a set of proteins, compared with LTBI patients (Figure 2B, right), including several basic proline-rich proteins; cystatins S, D and N; carbonic anhydrase 6; and secreted-frizzle related protein 1. String 10.1 analysis (Figure 3C) indicated that these proteins play a role in bitter-taste perception and endopeptidase activity (Supplementary information).

Finally, to select proteins that are specifically increased in LTBI patients, we crosschecked the list of proteins that were accumulated in LTBI versus active TB patients (Figure 2A, left) with the list of proteins that were accumulated in LTBI versus uninfected contacts (Figure 2B, left). Only five proteins, mammaglobin-B, retinal dehydrogenase 1, ectopic-p-granules protein 5 homolog and BPI-fold-containing family A and B members, are accumulated in LTBI versus both uninfected and active TB patients (Figure 3D). In this case, due to the low number of proteins, String 10.1 pathway analysis was not possible.
The ratios for all statistically significant proteins are represented as box-plots (supplementary info and Figures 4 and 5). Proteins such as alpha-1-acid glycoproteins 1 and 2, haptoglobin, fibrinogen alpha and beta and protein S100P were increased in active TB versus both LTBI and uninfected contacts (Figure 4). This result indicates that the increase in these proteins is characteristic of active TB patients, which differentiates active TB patients from uninfected and LTBI contacts. Similarly, proteins that were detected to be increased in uninfected contacts versus LTBI patients, such as carbonic anhydrase 6, mucin-7, cystatin S, secreted frizzle-related protein 1, basic salivary protein 2 and cysteine-rich secretory protein 3 (Figure 5), were also increased in uninfected versus active TB patients.

3.3. **Specific protein signature of active TB patients and contacts in saliva**

A similar statistical analysis was followed to select the specific proteomic signature in saliva samples. The volcano plots (Figure 6A, right) suggest an accumulation of a subset of proteins in the saliva of active TB patients versus LTBI patients, such as haptoglobin, fibrinogens and alpha-1-acid glycoprotein 1 and 2. String analysis demonstrated a strong interaction network (Figure 7A) between proteins mainly related to complement activation and acute-phase response Supplementary information 3).

In this case, the proteins that were decreased in the saliva of active TB versus LTBI patients (Figure 6A, left) were demonstrated to share a very strong interaction network (Figure 7B) and were related to carbohydrate metabolism and GTPase-mediated signal transduction (Supplementary info 3).

Unlike the sputum, the saliva samples of uninfected and LTBI contacts did not exhibit significant differences (Figure 6B). Only mucin-like protein 1 appeared to be slightly accumulated in uninfected contacts (Figure 6B, right). On the other hand,
cathepsin G and cystatin-F (Figure 6A, left) were detected to be slightly increased in the saliva of LTBI patients versus both uninfected and active TB patients (Figure 7C).

Box-plots representing the three ratios were prepared for all the significant proteins in the saliva (Proteomic data repository and Figures 8 and 9). As observed in the sputum samples, several proteins including haptoglobin, alpha-1-acid glycoprotein 1 and 2, immunoglobulin-gamma 4 chain, fibrinogens and dermcidin (Figure 8) were specifically increased in active TB patients versus both uninfected and LTBI contacts. Another set of proteins, which is represented by glutathione synthetase, lactoylglutathione lyase, protein disulfide isomerase, triose-phosphate isomerase, tropomyosin alpha 4 and ras GTPase-activating-like protein (Figure 9), was specifically decreased in active TB patients versus both uninfected and LTBI contacts.

3. Discussion

The sputum and saliva have been, classically, a source of information for the study of Mycobacteria (23) and have been used, more recently, for detection of the pathogen *M. tuberculosis* (24). Separately, these two fluids have been used for biomarker discovery in different diseases (25, 26); however, only few reports compare both fluids (27, 28). The present study is, to our knowledge, the first to focus specifically on biomarker discovery in TB through combining both types of samples.

*Differences between saliva and sputum in terms of protein composition*

Analysis of the modulation of the identified proteins indicated strong differences between the fluids. Global quantification results (Supplementary figure 1 A) suggested that the sputum is, as expected, a more complex biological fluid than saliva. Dispersion diagrams (Supplementary figure 1B, C and D) show that no correlation exists when
comparing quantification ratios from the saliva versus the sputum samples for all the possible ratios (uninfected/LTBI, uninfected/TB and LTBI/TB), which indicates that the saliva and sputum differ in protein composition and proportion.

Specific proteomic signature of active TB patients

Statistical analysis of the quantification ratios for the two biological fluids demonstrated not only similarities but also some important differences. Proteins related to cell degranulation, inflammation, acute-phase response and defense against bacteria appeared to be accumulated in the sputum and saliva of the active TB patients (Figures 2A and 6A) compared to that in the uninfected and LTBI contacts (Figures 4 and 8). This is in concordance with previous studies on the serum (29) and saliva (30) using antibody-based techniques and on the serum (31) using MS-based technology.

Additionally, the active TB patients presented, specifically in the saliva, a very significant decrease in the proteins related to sugar metabolism and GTPase-related signal transduction. (Figure 6A, left). Specifically, changes in proteins related to carbohydrate metabolism such as those represented in Figure 9 concur with previous studies. Shin et al., detected an imbalance in carbohydrate and lipid metabolism using NMR-based metabolomics in the tissue of mice that were infected with MTB (32). More recently, using similar approaches, Zhou et al. reported an increase in anaerobic glycolysis rate in the sera of TB patients (33), which was characterized by elevated levels of lactate and pyruvate. Our findings may represent a reflection of the disease-associated metabolome adaptations of both the microbe and host, as has been previously described using metabolomics (34, 35). Since the adaptive immune response systems rely on various microbiota interactions to promote immune cell maturation and function
(36), a microbial imbalance due to active TB infection may result in a weakened immune system and a loss of ability to fight disease.

**Specific proteomic signature in uninfected contacts**

In contrast, the uninfected contacts presented a proteomic specific signature only in the sputum (Figure 2B, right) and not in the saliva (Figure 6B, right), where no significant differences were found between the uninfected and LTBI contacts. In the sputum, an increase was detected for a subset of proteins that play a role in endopeptidase activity and bitter-taste perception in the uninfected contacts versus both LTBI and active TB patients (Figure 5).

Bitter-taste perception has been revealed in recent years as a key regulator of the innate immune system in the respiratory tract (37-39). The central core components of this machinery are the type-2 receptors (T2Rs) that belong to the G-protein-coupled receptor family (GPCRs). Humans are known to have at least 25 different T2R subtypes that are expressed in several cell types in the airways (40). Activation of T2Rs by secreted bacterial substances promotes, via calcium-dependent signaling, the secretion of antimicrobial peptides in the respiratory tract during acute bacterial infection (41). On the other hand, many regulators of the anti-inflammatory process during infection and allergy exhibit endo-peptidase activity, especially serine and cysteine protease inhibitors (42). In recent years, new therapeutic approaches have been based on the use of serine protease inhibitors (43, 44).

Since the protein content of the sputum is a reflection of the proteins secreted by the oral and nasal mucosa, we postulate that these proteins could be conferring a special innate immune status that protects these individuals against MTB infection, which helps to eliminate the pathogen before it reaches the lungs. Some of the proteins that were detected to be increased in the sputum of the uninfected contacts have been previously
identified in the oral epithelia (45), and they play a role in the defense against other pathogenic bacteria. Cystatin-S and carbonic anhydrase 6 are important players in the bitter-taste perception machinery (46). This process has emerged in recent years as a key regulator of the innate immune response. In a CAH6 null mouse model presents alterations in the lower respiratory tract reduced defense capacity and the renewal of the lining epithelium (46, 47). Mucin-7, a glycosylated member of the mucin family (48), binds selectively to *Staphylococcus aureus* in the respiratory tract (49). Mucins are produced mainly by the goblet cells of the respiratory tract and are considered as highly abundant proteins in the oral and nasal fluids.

Specific proteomic signature of LTBI patients

The discovery of new biomarkers for latent TB infection is of paramount importance to accomplish the goals of the WHO "End TB" strategy. This asymptomatic condition could be prolonged for many years or even the lifetime, and it still lacks a sensitive and specific method for diagnosis. The saliva and sputum are, a priori, ideal fluids for this search due to the ease and lack of invasiveness in sample collection. A bona fide latent TB infection biomarker should be specific to LTBI patients and should differentiate them from both uninfected and active TB patients.

In the present study, we have focused on proteins that were specifically increased in LTBI patients, five in the sputum (Figure 3D) and only two in the saliva (Figure 7C). In the sputum, ectopic p-granules 5 homolog (EPG5), retinal dehydrogenase 1 and mammaglobin B are accumulated in LTBI patients versus uninfected and TB patients (Figure 10). EPG5 has been revealed to be part of the autolysosomal formation machinery (50). Mutations in its coding gene are the causal agent of Vici syndrome, a recessive genetic condition that is characterized by
immunodeficiency. Retinal dehydrogenase 1 is involved in detoxification of lipid
aldehydes (51), and mammaglobin-B, a member of the uteroglobin family, is a steroid-
binding protein (52). In the saliva, we found that cathepsin-G and cystatin-F were
specifically accumulated in LTBI patients. Cathepsin-G is an antibacterial protein with
activity against Gram-negative bacteria (53), and cystatin-F has been recently identified
as a key factor in the differentiation process from monocytes to macrophages (54, 55).

Considerations on the biological variability of the proteomic results

All the proteins selected present significant (p-value ≤ 0.0001) differences in the
quantification ratios TB/LTBI, uninfected/LTBI and uninfected/TB. However, individual box-plot diagrams show, in some cases, large variability in the quantification
ratios, which result eventually in "outliers". Regarding this observation, it is important
to consider two possibilities: a) the presence of "undetectable" LTBI patients in the
uninfected contact group and b) the presence of nondiagnosed active TB patients in the
LTBI patient group. Although the criteria followed for the diagnosis and classification
of the patients in the EMI-TB cohort has followed the highest clinical standards, we
cannot totally discard the possibility of having enrolled eventual false uninfected
contacts or undiagnosed active TB patients.

4. Conclusion

In summary, our work represents a step forward in studying the mechanisms
triggered in the host by MTB infection using shotgun proteomics. Our findings that
indicate accumulation of acute-phase response and inflammation players in active TB
patients agree with previous targeted, antibody-based studies. The observation that
several enzymes involved in sugar metabolism were decreased in these patients is in
concordance with recent metabolomic studies, which indicate an imbalance in the glycolytic rate of the host during active TB. We also found a small set of proteins that were specifically increased in latent TB patients. In this case, further studies and larger patient cohorts are now required to decipher the biological and mechanistic significance of this finding.

Finally, proteins that were identified to be related to the innate immune response were significantly overrepresented in the sputum of uninfected individuals who have been in close contact with an active TB patient; this result suggests that nasal and oral mucosa play a critical role in the initial entry of the pathogen. This opens a new window of opportunity for modulating their expression with the use of specific adjuvants, thus enhancing the innate immune response as the first barrier against infection. Furthermore, it supports the importance of nasal vaccination to fight TB. Additionally, since the identified proteins play biological roles in nonclassical immune processes such as bitter-taste perception and endopeptidase activity, our results indicate that other biological and not necessarily immunity-related processes might play a role in the MTB infection or even in the reactivation of the disease. In our opinion, the special resistance status of a specific individual to MTB infection could be determined for the summed action of a combination of several “markers of resistance” rather than the action of only one “master protective player”. To gain more biological and clinical significance, our findings must be further validated in a functional model of MTB infection.

ACKNOWLEDGMENTS

We are very grateful to the patients who participated in the present study for their altruistic donation of samples and to the specialized personnel of “Complexo Hospitalario Universitario de Pontevedra” (SERGAS) for sample collection and
processing. Additionally, we would like to thank Manuel Marcos (Proteomics Facility, University of Vigo, CACTI, Vigo, Spain) for shotgun data acquisition and Alexandre Alonso-Fernández (CSIC) for help with statistics. CSIC (M.C.), CINBIO (A.G.F.) and SGUL (R.R.) groups are part of the Eliciting Mucosal Immunity in Tuberculosis (EMI-TB) consortium. EMI-TB is a European Union Horizon 2020 funded action (Grant No 643558) that is focused on selecting and developing a novel vaccine candidate for tuberculosis (TB). We would like to thank all its members, especially, Elina Garet and Silvia Lorenzo-Abalde, for their collaboration and for helpful discussion and suggestions.
REFERENCES


**Table 1.** Demographic summary of the patients included in the EMI-TB cohort.

<table>
<thead>
<tr>
<th>Patient Type</th>
<th>Gender</th>
<th>Age (mean ±SD)</th>
<th>Contact score (mean ±SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active TB (n=26)</td>
<td>Female (15.4%), male (84.6%)</td>
<td>41.3 ± 13.9</td>
<td>n.a.</td>
</tr>
<tr>
<td>LTBI (n=29)</td>
<td>Female (41.3%), male (58.7%)</td>
<td>47.4 ± 14.7</td>
<td>10.9 ± 2.6</td>
</tr>
<tr>
<td>Uninfected (n=44)</td>
<td>Female (52.3%), male (47.7%)</td>
<td>40.0 ± 15.2</td>
<td>9.6 ± 2.3</td>
</tr>
</tbody>
</table>
Table 2. Individual samples used for the shotgun proteomic study.

<table>
<thead>
<tr>
<th>SAMPLE</th>
<th>TMT 1</th>
<th>TMT 2</th>
<th>TMT 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMT-Label</td>
<td>Group</td>
<td>SAMPLE</td>
<td>PO-20</td>
</tr>
<tr>
<td>126</td>
<td>TB</td>
<td>PO-20</td>
<td>TB</td>
</tr>
<tr>
<td>127N</td>
<td>TB</td>
<td>PO-28</td>
<td>TB</td>
</tr>
<tr>
<td>127C</td>
<td>TB</td>
<td>PO-38</td>
<td>TB</td>
</tr>
<tr>
<td>128N</td>
<td>TB</td>
<td>PO-24</td>
<td>TB</td>
</tr>
<tr>
<td>128C</td>
<td>TB</td>
<td>PO-27</td>
<td>TB</td>
</tr>
<tr>
<td>129N</td>
<td>LTI</td>
<td>PO-30</td>
<td>LTI</td>
</tr>
<tr>
<td>129C</td>
<td>LTI</td>
<td>PO-42</td>
<td>LTI</td>
</tr>
<tr>
<td>128N</td>
<td>LTI</td>
<td>PO-46</td>
<td>LTI</td>
</tr>
<tr>
<td>128C</td>
<td>LTI</td>
<td>PO-22</td>
<td>LTI</td>
</tr>
<tr>
<td>129N</td>
<td>LTI</td>
<td>PO-52</td>
<td>LTI</td>
</tr>
<tr>
<td>129C</td>
<td>LTI</td>
<td>PO-62</td>
<td>LTI</td>
</tr>
<tr>
<td>130N</td>
<td>nonLTI</td>
<td>PO-34</td>
<td>nonLTI</td>
</tr>
<tr>
<td>130C</td>
<td>nonLTI</td>
<td>PO-54</td>
<td>nonLTI</td>
</tr>
<tr>
<td>131</td>
<td>nonLTI</td>
<td>PO-44</td>
<td>nonLTI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SAMPLE</th>
<th>TMT 1</th>
<th>TMT 2</th>
<th>TMT 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMT-Label</td>
<td>Group</td>
<td>SAMPLE</td>
<td>PO-20</td>
</tr>
<tr>
<td>126</td>
<td>TB</td>
<td>PO-20</td>
<td>TB</td>
</tr>
<tr>
<td>127N</td>
<td>TB</td>
<td>PO-28</td>
<td>TB</td>
</tr>
<tr>
<td>127C</td>
<td>TB</td>
<td>PO-38</td>
<td>TB</td>
</tr>
<tr>
<td>128N</td>
<td>TB</td>
<td>PO-24</td>
<td>TB</td>
</tr>
<tr>
<td>128C</td>
<td>TB</td>
<td>PO-27</td>
<td>TB</td>
</tr>
<tr>
<td>129N</td>
<td>LTI</td>
<td>PO-30</td>
<td>LTI</td>
</tr>
<tr>
<td>129C</td>
<td>LTI</td>
<td>PO-42</td>
<td>LTI</td>
</tr>
<tr>
<td>128N</td>
<td>LTI</td>
<td>PO-46</td>
<td>LTI</td>
</tr>
<tr>
<td>128C</td>
<td>LTI</td>
<td>PO-22</td>
<td>LTI</td>
</tr>
<tr>
<td>129N</td>
<td>LTI</td>
<td>PO-52</td>
<td>LTI</td>
</tr>
<tr>
<td>129C</td>
<td>LTI</td>
<td>PO-62</td>
<td>LTI</td>
</tr>
<tr>
<td>130N</td>
<td>nonLTI</td>
<td>PO-34</td>
<td>nonLTI</td>
</tr>
<tr>
<td>130C</td>
<td>nonLTI</td>
<td>PO-54</td>
<td>nonLTI</td>
</tr>
<tr>
<td>131</td>
<td>nonLTI</td>
<td>PO-44</td>
<td>nonLTI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SAMPLE</th>
<th>TMT 1</th>
<th>TMT 2</th>
<th>TMT 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMT-Label</td>
<td>Group</td>
<td>SAMPLE</td>
<td>PO-20</td>
</tr>
<tr>
<td>126</td>
<td>TB</td>
<td>PO-20</td>
<td>TB</td>
</tr>
<tr>
<td>127N</td>
<td>TB</td>
<td>PO-28</td>
<td>TB</td>
</tr>
<tr>
<td>127C</td>
<td>TB</td>
<td>PO-38</td>
<td>TB</td>
</tr>
<tr>
<td>128N</td>
<td>TB</td>
<td>PO-24</td>
<td>TB</td>
</tr>
<tr>
<td>128C</td>
<td>TB</td>
<td>PO-27</td>
<td>TB</td>
</tr>
<tr>
<td>129N</td>
<td>LTI</td>
<td>PO-30</td>
<td>LTI</td>
</tr>
<tr>
<td>129C</td>
<td>LTI</td>
<td>PO-42</td>
<td>LTI</td>
</tr>
<tr>
<td>128N</td>
<td>LTI</td>
<td>PO-46</td>
<td>LTI</td>
</tr>
<tr>
<td>128C</td>
<td>LTI</td>
<td>PO-22</td>
<td>LTI</td>
</tr>
<tr>
<td>129N</td>
<td>LTI</td>
<td>PO-52</td>
<td>LTI</td>
</tr>
<tr>
<td>129C</td>
<td>LTI</td>
<td>PO-62</td>
<td>LTI</td>
</tr>
<tr>
<td>130N</td>
<td>nonLTI</td>
<td>PO-34</td>
<td>nonLTI</td>
</tr>
<tr>
<td>130C</td>
<td>nonLTI</td>
<td>PO-54</td>
<td>nonLTI</td>
</tr>
<tr>
<td>131</td>
<td>nonLTI</td>
<td>PO-44</td>
<td>nonLTI</td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

**Fig 1:** Schematic design and workflow of the proteomic shotgun analysis.

**Fig 2:** Volcano-plot representations of the statistical analysis in sputum. Analysis of the active TB/LTBI (A) and uninfected/LTBI (B) ratios for all the 469 proteins quantified in the three TMTs with at least two unique peptides.

**Fig 3:** String 10.1 interaction pathway analysis of the specific protein signature of active TB patients (A and B), uninfected contacts (C) and LTBI patients (D) in sputum samples.

**Fig 4:** Box-plot representations of the quantification ratios of selected proteins accumulated in the sputum of active TB patients.

**Fig 5:** Box-plot representations of the quantification ratios of selected proteins accumulated in the sputum of uninfected contacts.

**Fig 6:** Volcano-plot representations of the statistical analysis in saliva. Analysis of the active TB/LTBI (A) and uninfected/LTBI (B) ratios for all the 379 proteins quantified in the three TMTs with at least two unique peptides.

**Fig 7:** String 10.1 interaction pathway analysis of the specific protein signature of active TB (A and B) and LTBI (C) patients in saliva samples.

**Fig 8:** Box-plot representations of the quantification ratios of selected proteins accumulated in the saliva of active TB patients.

**Fig 9:** Box-plot representations of the quantification ratios of selected proteins decreased in the saliva of active TB patients.

**Fig 10:** Box-plot representations of the quantification ratios of selected proteins accumulated in the sputum and saliva of LTBI patients.
Saliva and Sputum Samples
- Active TB patients (TB)
- Infected household contacts (LTBI)
- Uninfected household contacts (Uninfected)

Protein precipitation, quantification and checking of the integrity

Protein digestion and TMT10plex labelling
- 9 active TB patients
- 9 LTBI contacts
- 9 uninfected contacts
  for each biological fluid.

High pH Reversed-phase peptide fractioning

LC-Orbitrap-Elite analysis

ProteomeDiscoverer 2.1 identification and quantification

Sputum: 1218 proteins; Saliva: 847 proteins
  Quantified with at least two unique peptides

Sputum: 766 proteins; Saliva: 562 proteins
  Quantified in the three TMTs

Sputum: 469 proteins; Saliva: 379 proteins
  p-value ≤ 0.0001 after Kruskal-Wallis analysis
  of all the individual quantification ratios (n=81)
  and minimum 1.5 fold change

Proteomic Signature of active TB and contacts
Figure 2
Click here to download Figure: new Figure 2.pdf
A SPUTUM: High in TB vs. LTBI

B SPUTUM: Low in TB vs. LTBI

C SPUTUM: High in uninfected vs. LTBI

D SPUTUM: High in LTBI vs. TB/uninfected

Inflammation and acute-phase response

Figure 3
Click here to download Figure: new Figure 3.pdf
Figure 4
Click here to download Figure: new Figure 4.pdf
Figure 6

Click here to download Figure: new Figure 6.pdf

-log10 p-val

log2 fold change TB vs. LTBI

p-val < 0.0001

p-val > 0.0001

-log10 p-val

log2 fold change uninfected vs. LTBI

p-val < 0.0001

p-val > 0.0001

Figure 6

Click here to download Figure: new Figure 6.pdf
Inflammation and acute-phase response

**SALIVA:** High in TB vs. LTBI

**Figure 7**

Click here to download Figure: new Figure 7.pdf

GTPase-mediated signal transduction

**Monosaccharide metabolism**

**SALIVA:** Low in TB vs. LTBI

**SALIVA:** High in LTBI vs. TB/uninfected

**SALIVA:** High in TB vs. LTBI

---

- **MONOSACCHARIDE METABOLISM**
- **GTPase-mediated signal transduction**
- **INFLAMMATION AND ACUTE-PHASE RESPONSE**
Figure 8
Click here to download Figure: new Figure 8.pdf
Figure 9
Click here to download Figure: new Figure 9.pdf
Figure 10

Click here to download Figure: new Figure 10.pdf
Supplementary figure 1
Click here to download Supplementary material: new Supplementary figure 1.pdf
Click here to download Supplementary material: new Supplementary Info 2 Pathway analysis SPUTUM.xlsx
Supplementary info 3

Click here to download Supplementary material: new Supplementary Info 3 Pathway analysis SALIVA.xlsx
Conflict of Interest

Click here to download Conflict of Interest: Conflict of interest.docx