Evaluation of the potential of *Mycobacterium smegmatis* as vaccine Candidate against tuberculosis by *in silico* and *in vivo* studies

Le Thuy Nguyen Thi, Reinier Borrero Maura, Sonsire Fernandez, Giselle Reyes, José Luis Perez, Fátima Reyes, María de los Angeles García, Midrey Fariñas, Juan Francisco Infante, Yanely Tirado, Alina Puig, Gustavo Sierra, Nadine Álvarez, Juan Carlos Ramírez, María Elena Sarmiento, Mohd-Nor Norazmi, Armando Acosta*

1 Biotechnology Centre of Ho Chi Minh City, 176 Hai Ba Trung Street, Ho Chi Minh City, Vietnam.
3 School of Health Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Malaysia

Email: aracosta@finlay.edu.cu

In this study, we scanned multiple published databases of gene expression *in vivo* of *M. tuberculosis* at different phases of infection in animals and humans, to select 38 proteins that are highly expressed in the active, latent and reactivation phases. The selected proteins were predicted for T and B epitopes. For each proteins, the regions containing T and B epitopes were selected at the same time to look for identical epitopes on *M. smegmatis* based on sequence alignments. Preliminary studies of humoral immunogenicity and cross-reactivity with *M. tuberculosis* in mice using two *M. smegmatis*-derived experimental vaccines were carried out, demonstrating the immunogenicity of *M. smegmatis* proteoliposomes and the recognition of *M. tuberculosis* proteins by the sera of animals immunized with this vaccine candidate. The conjunction of *in silico* and *in vivo* studies suggested the potential for future evaluation of *M. smegmatis* as vaccine candidate against tuberculosis using different strategies.

**Keywords:** *Mycobacterium tuberculosis*, *Mycobacterium smegmatis*, epitopes, prediction.

Introduction

*Mycobacterium tuberculosis*, the main agent of tuberculosis (TB) in humans caused 9.2 million new cases and 1.7 million deaths in 2006. It is estimated that up to one third of the world’s population has been infected with *M. tuberculosis* (1). The current vaccine, *Mycobacterium bovis* Bacillus Calmette–Guerin (BCG) is the most widely used in humans. However, the uncertain efficacy of BCG vaccine in the prevention of pulmonary TB in adults, and the emergence of extensively drug resistant *M. tuberculosis* strains have further enhanced the sense of urgency in the development of new TB vaccines (2). With the complete genome sequence of *M. tuberculosis* and the advances in the sequencing of BCG and other mycobacteria, an enormous flow of valuable information has been made available for new vaccine strategies (2, 3).

Vaccine design is highly suited to the application of *in silico* techniques, for both discovery and development of new and existing vaccines (3). Bioinformatic tools allow selecting regions of microbial genomes that trigger immune responses against vaccine components. In order to give a bigger impact to the bioinformatic studies it is important to take into consideration the data obtained from studies of gene expression *in vivo* of *M. tuberculosis*. This combination will focus the studies on a smaller group of relevant antigens to be used in vaccine development (3).

The fast-growing and nonpathogenic soil bacterium *Mycobacterium smegmatis* which shares many features and identical genomic sequences with the pathogenic *M. tuberculosis* (4) is a potential candidate for developing new TB vaccines. The use of *M. smegmatis* is expected to have more advantages for TB vaccine development, feasible for research in normal laboratories, saving time and cost. With the aim to evaluate the antigenic similarity between *M. smegmatis* and *M. tuberculosis*, a combination of *in silico* and *in vivo* studies was carried out.

Materials and methods

Identification and selection of *in vivo* expressed genes of *M. tuberculosis*

A search using the Internet Google Search Tool to find articles reporting *in vivo* studies of *M. tuberculosis* gene expression in humans and animals was carried out. The genes that are highly expressed in common in humans and animals at all stages of infection: activation, latency and reactivation were selected for epitope prediction.

---

* Dr. en Ciencias Médicas y Médico Especialista de 2do. Grado en Inmunología, Investigador Titular y Profesor Titular. Director del Dpto.de Biología Molecular de la Vicepresidencia de Investigaciones del Instituto Finlay.
Epitope prediction

The amino acid sequences and the features of the selected genes were obtained from the Reference Sequence (RefSeq) database (http://www.ncbi.nlm.nih.gov/RefSeq/) at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/).

T-cell epitope prediction

HLAPred server was used for prediction of T-cell epitopes in this study. The program is developed by Bioinformatics Centre Institute of Microbial Technology, India (http://www.imtech.res.in/raghava/hlapred/). All the alleles from the list of 36 HLA class I and 51 HLA class II were selected for the prediction.

B-cell epitope prediction

Two servers, Bcepred and ABCpred, were combined for linear B-cell epitope prediction. Bcepred server is accessible at http://www.imtech.res.in/raghava/bcepred/ and ABCpred web server is accessible at www.imtech.res.in/raghava/abcpred/. For Bcepred server we combined of 7 physico-chemical properties of amino acids (hydrophilicity, flexibility, accessibility, polarity, exposed surface and turns and antigenic propensity). For ABCpred server we predicted B epitope length of 16 amino acids.

Searching for identical epitopes on M. smegmatis

The regions in the sequences of the 38 selected proteins of M. tuberculosis containing T and B epitopes at the same time were used to identify the identical epitopes on M. smegmatis genome using BLAST.

(Please provide the URL for the BLAST search)

Prediction of subcellular localization of proteins from M. tuberculosis and M. smegmatis

The subcellular localization of the selected M. tuberculosis proteins were defined according to the report of the identification and localization of 1044 M. tuberculosis proteins using two-dimensional, capillary high-performance liquid chromatography coupled with mass spectrometry (2DLC/MS) method (5).

The amino acid sequences of M. smegmatis proteins containing identical epitopes with M. tuberculosis and their features were obtained from the Reference Sequence (RefSeq) database (http://www.ncbi.nlm.nih.gov/RefSeq/) at NCBI. To determine the subcellular location of the proteins of M. smegmatis and the proteins of M. tuberculosis that were not in the first list of 1044 proteins, we used TBpred server. TBpred is freely available at http://www.imtech.res.in/raghava/tbpred/

Evaluation of the immunogenicity and the cross-reactivity with M. tuberculosis on mice immunized with M. smegmatis vaccine candidates

M. smegmatis culture

Lyophilized cells of M. smegmatis from Molecular Biology Laboratory, Finlay Institute, were used in this study. The cells were grown at 37 °C, shaken at 200 rpm in culture broth medium containing glycerol 0.05 % and 0.05 % tween 80.

Preparing proteoliposomes from M. smegmatis

Biomass of M. smegmatis was obtained from centrifugation of the broth culture at 17700 x g (centrifuge J2-21, Beckman, EUA) during 1 hour, at 4 °C. The biomass was suspended in buffer (10-30 mM Tris, 2-10mM EDTA, pH 7-8.5) and washed twice. After suspension of the pellet in the same buffer, sodium deoxycholate (5-15%) was added with the ratio 0.1-0.25 mL/g biomass. After centrifugation with the same conditions, the supernatant was ultra-centrifugated at 50000-70000 x g (centrifuge L7-65 Beckman, EUA) for 2-8 hours at 4 °C. The sediment was resuspended in the buffer and filtrated with Sartorius Minisart-plus filter 0.45-0,2 µm.

Mice immunization

Fifteen female Balb/c mice of 6-8 weeks of age (16-18 g weight) obtained from the National Center of Laboratory Animals (CENPALAB), La Habana, Cuba, were used. The animals were housed in appropriate biological containment facilities at Animal Laboratory, Finlay Institute. All the experimental procedures were carried out following the guide for the use of Laboratory Animals (EEC Council Directive, 1987). The animals were distributed into 3 experimental groups (n=5): (i) negative control group consisted of mice receiving one dose at T0 of PBS; (ii) mice receiving two doses of proteoliposomes of M. smegmatis (50 µg/dose/mouse) in Freund’s incomplete adjuvant at T0 and T21 (day 21 after the first immunization); (iii) mice immunized by a single dose at T0 with live cells of M. smegmatis (1 x 10⁶ cfu).

Mice of all groups were injected via the subcutaneous route. 42 days after the first inoculum, the blood of mice of all groups was collected and centrifugated to obtain the sera.

ELISA to determine antibody responses

Flat bottomed polystyrene ELISA plates (Maxisorp, Nunc) were coated (100 µL/well) 16 hours at 4 °C with 5 µg/mL M. smegmatis proteoliposomes, 5 µg/mL whole cell lysate of M. tuberculosis. 10⁶ cfu/mL of live-cell M. smegmatis were coated 16 hours at 37 °C. The antigens were dissolved in the coating buffer (Carbonate/Bicarbonate 0.1 M, pH 9.6) and added to the plates. The plates were blocked with blocking buffer (PBS...
with 3% skim milk) for 1 hour at 37 °C. Serum samples were diluted 1/50 with 1% skim milk (w/v) in PBS (NaCl 137 mM, KCl 2.7 mM, NaHPO4·7H2O, KH2PO4 1.4 mM, pH 7.4 and were added (100 µL/well) into each well. The samples were assayed in duplicate and were incubated 1 hour at 37 °C. The plates were incubated for 1 hour at 37 °C with the peroxidase-conjugated secondary antibody (affinity-purified anti-mouse IgG; Sigma, cod A-4416) in a dilution of 1/1000 (100 µL/well). The reaction was developed with the solution of o-phenylenediamine (OPD; Sigma): 5 mg of OPD in 12.5 mL of sodium citrate buffer 0.1M, pH 5, H2O2 5 µL (100 µL/well). The reaction was stopped with 50 µL/well of H2SO4 2N after 20-30 minutes incubation with OPD solution. The absorbance was recorded at 492 nm with an ELISA plate reader (Multiskan Multisoft). Each experiment was carried out twice.

Statistical analysis

ELISA results were analyzed using ANOVA to determine significant differences among the groups at the 95% confidence level. Multiple range test were used to determine pairs with significant differences.

Results

Identification and selection of in-vivo highly expressed genes of M. tuberculosis

In this study, we analyzed published articles of gene expression of M. tuberculosis during different phases of infection in animals and humans (6-24): from humans, animals (mainly on mice and rabbits) and from both humans and mice. These studies were directed to identify M. tuberculosis genes of responsible for the adaptation and evasion from the host immune responses and drugs as well as for the reactivation of infection and invasion of the bacteria. The expression profile under different conditions of 3924 genes from M. tuberculosis during different stages of infection in different hosts was identified (11, 12, 16-19).

A total of 38 M. tuberculosis genes that were reported significantly up-regulated at the active, latent and reactivation phases were selected from the whole genome. The amino acid sequences obtained and the features of the selected genes from the Reference Sequence (RefSeq) database the subcellular localization of the corresponding proteins were identified.

Subcellular localization prediction provides important information about a protein that allows making inferences regarding to the protein’s function, to annotate genomes, and particularly in the case of bacterial pathogens to identify potential diagnostic, drug and vaccine targets. The distribution of subcellular location of 38 selected proteins are shown in Figure 1.

Searching for identical epitopes on M. smegmatis

The regions containing T and B epitopes at the same time of 38 selected M. tuberculosis proteins were the bases to look for identical epitopes on M. smegmatis. In these regions, 1367 T-cell epitopes and 140 B-cell epitopes were identified on M. tuberculosis. Using BLAST a search was made on the whole genome of M. smegmatis. 68 identical T-cell epitopes and 11 identical B-cell epitopes on 18 proteins of M. smegmatis were recorded from the alignments.

From the prediction, 5 proteins are secreted, 8 proteins are located in cytoplasm and 5 proteins locate in all of the cell compartments of the bacteria.

It is interesting to note that all the 18 identified M. smegmatis proteins containing the identical epitopes with M. tuberculosis have assigned functions reported in the Reference Sequence database at NCBI that correlate with the functions of the 13 M. tuberculosis identified target-proteins. This finding is consistent with the close similarity between M. smegmatis and M. tuberculosis at the genomic and functional levels. This similarity is the bases of the use of M. smegmatis as a model for the studies of the biology of M. tuberculosis and for the development of drugs against tuberculosis.

Regarding the proteins of M. tuberculosis containing common epitopes with M. smegmatis, the distribution according to the functional categories of the 13 identified target-proteins is shown in Figure 2. These 13 proteins mainly relate to the intermediary metabolism and respiration; virulence, detoxification, adaptation; lipid metabolism; cell wall and cell processes; conserved hypothetical functions of tubercle bacilli that were demonstrated highly expressed and important for the bacteria to adapt to different host conditions during different stages of infection (11, 12, 16-19, 21, 22).

Figure 1. A Venn diagram of the subcellular distribution of 38 identified proteins of M. tuberculosis.
Classification of *M. tuberculosis* proteins according to functional categories. The distribution according to functional categories of the selected 38 highly expressed genes in vivo and of the subgroup of these genes sharing identical epitopes with proteins of *M. smegmatis* (13 target-proteins) is shown. The distribution of genes of *M. tuberculosis* whole genome is shown as reference.

**Evaluation of the immunogenicity of *M. smegmatis* and the cross-reactivity with *M. tuberculosis* on immunized mice**

*In vivo* studies to evaluate the immunogenicity and the cross-reactivity of two vaccine candidates from *M. smegmatis*: proteoliposomes (PL) and live cells were carried out by an indirect ELISA to determine the recognition of *M. smegmatis* and *M. tuberculosis* antigens by IgG antibodies of mice immunized with these two experimental vaccines. The results are shown in Figure 3.

Regarding the immunogenicity study, there was a statistical increase of the specific antibodies in the group of mice immunized with PL of *M. smegmatis* against whole cells of *M. smegmatis* and whole cell lysate of *M. tuberculosis* compared with the other groups (Figure 3). In the case of the responses against PL of *M. smegmatis*, animals immunized with PL had statistical increases of antibodies compared with the negative control group (mice immunized with PBS) (Figure 3). These results showed a better immunogenicity of PL with the adjuvant and immunization schedule used.

In the case of mice immunized with live cells of *M. smegmatis*, there were not statistically significant differences with negative control group against all the antigens tested. In this case, whole cells or PL of *M. smegmatis* and whole cell lysate of *M. tuberculosis* were not recognized by the antibodies of mice immunized with live cells of *M. smegmatis*. It should be taken into account that only one dose of immunization was used (as usually used of BCG vaccination). In a study using *M. smegmatis* as carrier vector for the expression of HIV proteins a more intensive immunization schedule was used with repeated immunizations in consecutive days (4).

**Discussion**

The expression of different genes during *in vivo* infection is very important for the rational design of vaccines, considering that only antigens expressed *in vivo* will have an effective interaction with the immune system and enable to elicit protective immune responses, so the study of this group of antigens should have the priority for the identification of relevant antigens/epitopes to be included in new vaccines.

One of the main objectives of this study was to identify *M. tuberculosis* genes that are highly expressed *in vivo* in order to establish a pre-selected pool of antigens for the search of potential protective epitopes. In our case, we searched for genes expressed *in vivo* in humans and animals and in all phases of infection.

This is important in all strategies considering the need to have the correlation between the animal model and human, and on the other hand to have the inclusion of epitopes able to be operative in all the stages of the infection where
vaccination could be used (primary infection, latency and clinical infection).

The distribution of the 38 highly expressed genes found in this study is relevant for the development of vaccine candidates targeting these antigens because they represent proteins of high importance for the survival of the bacteria during different phases of infection. Consequently, the immune responses elicited against these proteins should have a detrimental effect upon the bacteria, thus contributing to the protection. On the other side, the high representation of the cell wall component proteins gives the opportunity to stimulate specific humoral immune responses. In this respect, the potential of humoral antibodies in the protection against mycobacterial infections was demonstrated (31).

The potential advantages of *M. smegmatis* as carrier or experimental vaccine against tuberculosis is related to the fact that it is rapidly destroyed by phagolysosomal proteases in the phagosomes of infected cells (32-34), which is different in other mycobacterial strains such as BCG that survive in host cells for months by inhibiting phagosome maturation. Nevertheless, *M. smegmatis* can induce cytokine production by macrophages better than pathogenic mycobacterial species and can activate and induce the maturation of dendritic cells better than BCG by upregulation of major histocompatibility complex (MHC) class I and costimulatory molecules. *M. smegmatis* can also access the MHC class I pathway for presentation of mycobacterial antigens more efficiently than BCG (4).

The recognition of *M. tuberculosis* whole cell lysate by animals immunized with PL of *M. smegmatis* demonstrated the humoral cross-reactivity between *M. smegmatis* and *M. tuberculosis* which matches the results obtained by *in silico* studies where five identical B-cell epitopes were identified from secreted, cell wall and membrane proteins of *M. smegmatis* compared with *M. tuberculosis*.

It should be considered that, in addition to the linear epitopes identified by the prediction algorithm, there could be identical conformational B-cell epitopes between *M. smegmatis* and *M. tuberculosis* reinforcing the humoral cross-reactivity. Beside, we could detect the humoral reactivity against lipid and carbohydrate components that are present in the vaccine candidates used in this study and in the different preparations used for the coating of the ELISA systems (PL, whole *M. smegmatis* and whole cell lysate of *M. tuberculosis*). In recent studies carried out in our laboratory (29) and in collaboration with NIBSC, UK (unpublished results) we demonstrated the humoral immunogenicity of mycobacterial lipidic components in humans. In the case of *M. smegmatis* it was revealed by the cryo-electron microscope that the outermost layer of...
**M. smegmatis** is an outer membrane with a bilayer structure of lipids (30). These components could be recognized by the humoral immune responses.

In the in vivo study, it was only studied the humoral cross-reactivity between **M. smegmatis** and **M. tuberculosis**, so it could not be ruled out the presence of cross-reactivity at cellular level after the immunization with the **M. smegmatis** vaccine candidates, a fact that should be explored in future studies taking into account the relevant role assigned to the cellular immune response in the protection against tuberculosis.

Studies carried out in our laboratory have demonstrated the recognition of proteins of **M. smegmatis** by IgG antibodies from pulmonary tuberculosis patients in a whole cell ELISA assay. In those studies the levels of specific antibodies in tuberculosis patients had a statistical increase compared with healthy controls (unpublished results). The result demonstrated that during in-vivo infection in humans **M. tuberculosis** expresses epitopes present in **M. smegmatis**, which matches the in silico and in vivo results obtained in current study, in addition it supports the idea of the potential use of **M. smegmatis** for the development of TB vaccine candidates.

**Acknowledgements**

This work was partially supported by the MOSTI Grant No:07-01-05-MEB007.

**References**

En este estudio se revisaron múltiples bases de datos publicadas, relacionadas con experimentos de expresión de genes de M. tuberculosis in vivo en diferentes estadios de la infección en humanos y animales. Se identificaron 38 proteínas con elevada expresión en las fases activa, latente y de reactivación de la infección. Se llevó a cabo la predicción de epítopes T y B en dichas proteínas. Las regiones de cada proteína que contenían simultáneamente epítopes T y B se seleccionaron y utilizaron para identificar regiones idénticas en M. smegmatis mediante el alineamiento de secuencias. Se llevaron a cabo estudios de inmunogenicidad humoral y reactividad cruzada con M. tuberculosis en ratones inmunizados con dos vacunas experimentales obtenidas a partir de M. smegmatis, demostrándose la inmunogenicidad de los proteoliposomas y el reconocimiento de proteínas de M. tuberculosis por el suero de ratones vacunados con este candidato vacunal. Los resultados obtenidos con los estudios in silico e in vivo sugieren la potencialidad para evaluación futura de candidatos vacunales obtenidos a partir de M. smegmatis para la prevención de la tuberculosis.

Palabras clave: Mycobacterium tuberculosis, Mycobacterium smegmatis, epítopes, predicción.