

Research Article

Identification of Novel Potential Vaccine Candidates against Tuberculosis Based on Reverse Vaccinology

Gloria P. Monterrubio-López, Jorge A. González-Y-Merchand, and Rosa María Ribas-Aparicio

Departamento de Microbiología, Escuela Nacional de Ciencias Biológicas (ENCB), Instituto Politécnico Nacional (IPN), Prolongación de Carpio y Plan de Ayala S/N, Colonia Santo Tomás, 11340 México, DF, Mexico

Correspondence should be addressed to Rosa María Ribas-Aparicio; rribas233@yahoo.com

Received 27 September 2014; Accepted 7 January 2015

Academic Editor: Tao Huang

Copyright © 2015 Gloria P. Monterrubio-López et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Tuberculosis (TB) is a chronic infectious disease, considered as the second leading cause of death worldwide, caused by *Mycobacterium tuberculosis*. The limited efficacy of the bacillus Calmette-Guérin (BCG) vaccine against pulmonary TB and the emergence of multidrug-resistant TB warrants the need for more efficacious vaccines. Reverse vaccinology uses the entire proteome of a pathogen to select the best vaccine antigens by *in silico* approaches. *M. tuberculosis* H37Rv proteome was analyzed with NERVE (New Enhanced Reverse Vaccinology Environment) prediction software to identify potential vaccine targets; these 331 proteins were further analyzed with VaxiJen for the determination of their antigenicity value. Only candidates with values ≥ 0.5 of antigenicity and 50% of adhesin probability and without homology with human proteins or transmembrane regions were selected, resulting in 73 antigens. These proteins were grouped by families in seven groups and analyzed by amino acid sequence alignments, selecting 16 representative proteins. For each candidate, a search of the literature and protein analysis with different bioinformatics tools, as well as a simulation of the immune response, was conducted. Finally, we selected six novel vaccine candidates, EsxL, PE26, PPE65, PE_PGRS49, PBPI, and Erp, from *M. tuberculosis* that can be used to improve or design new TB vaccines.

1. Introduction

Tuberculosis (TB) is a chronic infectious disease caused by an acid-fast bacillus, *Mycobacterium tuberculosis* [1]. TB is the second cause of death caused by an infectious agent throughout the world [2, 3]; in 2012, there were an estimated 8.6 million incident cases of TB globally, which is equivalent to 122 cases per 100,000 people, and the absolute number of cases continues to increase slightly from year to year [4].

The current vaccine against tuberculosis, bacillus Calmette-Guérin (BCG), exerts different levels of protection: from 46 to 100% against the disseminated disease form and from 0 to 80% against pulmonary disease [5, 6]. In addition to this low efficacy, reemergence of the disease caused by the appearance of the acquired immunodeficiency syndrome (AIDS) and multidrug-resistant (MDR) strains has generated requirements for a new and more efficient vaccine against TB [7].

The development of new vaccines starts with the identification of unique components of the microorganism capable of generating a protective immune response [3]. With traditional techniques, this could be a long and arduous process, aside from the difficulty of cultivating the microorganism in the laboratory [8–10].

Advances in sequencing technology and bioinformatics have resulted in an exponential growth of genome sequence information that has contributed to the development of software that aids genomic analysis in a short period of time and at a low cost. Reverse vaccinology (RV) applied to the genome of a pathogen aims to identify *in silico* the complete repertoire of immunogenic antigens that an organism is capable of expressing without the need of culturing the microorganism. Additionally, RV can help to discover novel antigens that might be less abundant, not expressed *in vitro*, or less immunogenic during infection that are likely to be missed by conventional approaches [8, 11–14].

The RV process begins with the proteomic information in a database; then, the selection of vaccine candidates is performed by means of different bioinformatics tools that analyze the properties of each protein and the human immune response generated by them [8–10, 15]. Good vaccine candidates are considered those that do not present homology with human proteins to avoid the generation of a potential autoimmune response [16]; these candidates must also lack transmembrane regions, in order to facilitate their expression. In addition, it is necessary to analyze the lack of cross-reaction among other pathogenic antigens [14]. Another characteristic a good vaccine candidate should have is to possess good antigenic and adhesin properties, which are important for the pathogenesis of the microorganism and for protection against the disease [13, 17]. Extracellular or cell surface localized proteins are good vaccine candidates due to their increased accessibility to the immune system [14, 16]. Currently, software useful for simulation of the immune response has been developed that could help in the search for novel vaccine candidates [15]. In this work, we have applied RV to the *M. tuberculosis* proteome with the purpose of selecting new antigens that could be used in a novel and more efficient vaccine against TB.

2. Materials and Methods

2.1. Proteome Analysis. New Enhanced Reverse Vaccinology (NERVE) software was downloaded, installed, and utilized to determine vaccine candidates employing the default parameters for Gram-positive bacteria [13]. The proteome sequences of *M. tuberculosis* H37Rv (NC_000962.2), *Mycobacterium bovis* AF2122/97 (NC_002945.3), and *M. bovis* BCG str. Pasteur I173P2 (NC_008769.1) were downloaded from the Genome Project database of the National Center for Biotechnology Information (NCBI) [18]. Each proteome was analyzed individually by NERVE; conservation values for all proteins were determined comparing the *M. bovis* and BCG proteome against the *M. tuberculosis* proteome using the comparative option.

2.2. Antigenicity Determination. The antigenicity value was calculated for each protein using its amino acid sequences and the VaxiJen server, which predicts whether a protein could be a protective antigen. VaxiJen is based on auto cross covariance (ACC) and has a threshold of 0.5 in the antigenicity value [19].

2.3. Selection of Representative Proteins. With the parameters calculated with NERVE and VaxiJen, we selected proteins that presented an antigenicity value ≥ 0.5 , 50% adhesin probability, and without homology with human proteins or transmembrane regions. The proteins selected were grouped according to the family of proteins to which they belong. In this manner, we obtained seven groups: ESX family proteins, PPE family proteins, PE family proteins, PE.PGRS family proteins, lipoproteins, hypothetical proteins, and, the last group, denominated “others,” composed of proteins with different miscellaneous characteristics. The amino acid sequence of each protein were downloaded from the NCBI protein data-

base, and an alignment was made for each group of proteins using Clustal X software [20] in order to select representative proteins from each group.

2.4. Immune Response Simulation. With the amino acid sequences of the proteins selected, a human immune response simulation was performed using the C-ImmSim software to predict whether these proteins could generate a protective immune response against TB [15]. C-ImmSim simulates a portion of a lymph node but is not set up to simulate a realistic concentration of antigen; however, we adjusted the antigen concentration simulation to a high dose, comparable to a vaccination event. Different immunizations were simulated with each protein in the following two different schemes: first, a single immunization with each protein individually at time zero and, second, three immunizations at 0, 2, and 4 weeks with each protein separately. The level of Th1 cells stimulated 80 days after the first injection was identified.

2.5. Protein Analysis. The bioinformatics programs used to study the vaccine candidate’s amino acid sequences included Phobius [21] to calculate and confirm protein subcellular localization more precisely, ANTHEPROT [22], Expasy [23], and IEDB software [24] and their different models for localizing protein regions with greater hydrophilic and greater solvent accessibility related with antigenic regions, and the SYF-PEITHI ver. 1.0 program [25], which was used to determine the frequency of presentation of peptides to 35 different alleles of the major histocompatibility complex (MHC). In the case of lipoproteins, we employed only ProPred software [26] to determine the frequency of presentation of 25 amino acid peptides to different alleles of the MHC-II.

2.6. Bibliographic Study. Bibliographic information was sought for each protein using different databases on the website for information regarding its putative function, its use as vaccines, its role in virulence, its corresponding evaluated mutants, its induction of an immune response, and its level of conservation in mycobacterial proteomes.

2.7. Vaccine Candidate Selection. The vaccine candidates were selected using all the results, simulations, and bibliographic information obtained. The candidates possess the best values of the parameters calculated and exert diverse functions that render them useful as different targets in the microorganism (Figure 1).

3. Results and Discussion

3.1. Proteome Analysis. RV offers the advantage of reducing the time and cost of the development process of a new vaccine with the advantage of being safer and more effective. With the purpose of designing a new vaccine against TB with a greater protection level against pulmonary disease, we utilized RV to select vaccine candidates from the *M. tuberculosis* proteome.

The selection of potential vaccine candidates in this study was based on the analysis of several important properties [13, 19, 21, 27]. (1) Surface proteins or secreted proteins were selected because they are good targets of the immune system

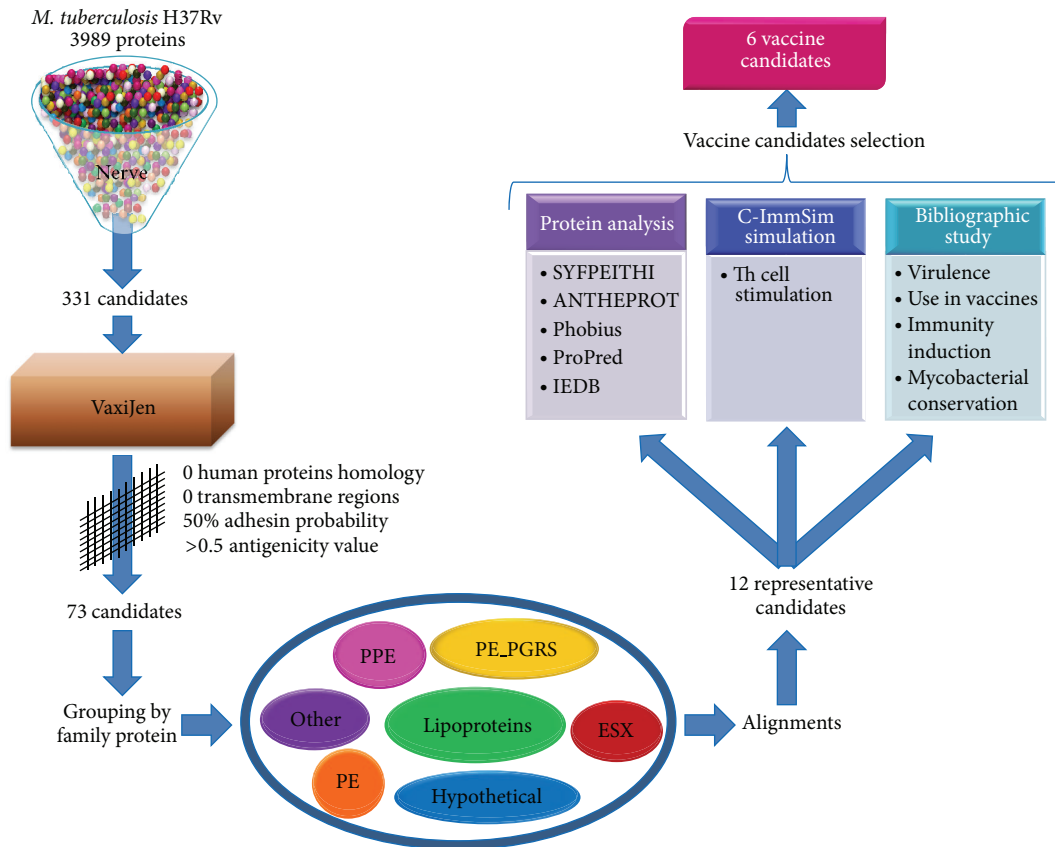


FIGURE 1: General workflow of the research. Reverse vaccinology was applied to the *M. tuberculosis* proteome to select novel vaccine candidates. The process starts with NERVE software selecting 331 vaccine candidates from 3989 proteins. These candidates were analyzed with different bioinformatics tools and bibliographic information selecting proteins representatives with the best values related with protective response. At the end of the study we chose six vaccine antigens (see details under Methods).

effector molecules. (2) Proteins with multiple transmembrane helices were discarded because they are not recommended for vaccine development, especially DNA vaccines, as they are difficult to clone, express, and purify. (3) Adhesin probability was considered an important factor since the first step in bacterial invasion is the contact with host molecules through adhesion structures, making adhesions good vaccine candidates capable of improving the immune response that results in blocking infection. (4) Proteins having similarity to those of the human proteome were avoided. The use of proteins or genes that encode them and having similitude with human proteins or DNA sequences can generate an autoimmune response or recombination and integration events in the host genome, respectively. (5) Proteins with the best values of antigenicity were chosen. Antigenicity is the property of the proteins to be recognized by the immune system; hence, it is desirable to find the highest antigenicity value for the selection of the best potential vaccine candidates. The *M. tuberculosis* proteome was studied with NERVE software, which identifies *in silico* vaccine candidates, analyzing the biological characteristics that influence vaccine design.

NERVE selected the *M. tuberculosis* H37Rv proteome, composed of 3989 proteins; the selection of candidates was performed considering the following characteristics: $\geq 50\%$

adhesin probability, fewer than two transmembrane regions, and fewer than five proteins similar to the human proteome. In addition to this, the candidates must lack either membrane or cytoplasmic localization. Finally, NERVE selected 331 proteins as vaccine candidates (Additional file 1) (see Supplementary Material available online at <http://dx.doi.org/10.1155/2015/483150>).

The results were compared with the information deposited in the VIOLIN database [28], and we found several important matches in some antigens. Those coincidences provided support for the results obtained with NERVE. The vaccine candidates selected have diverse putative functions and different conservation values. Moreover, this software tool has the option of comparing the proteomes of two different organisms and of determining a conservation value among all the proteins. In this case, we compared the *M. tuberculosis* H37Rv proteome against the *M. bovis* and the BCG proteome [13, 14].

The proteome analysis was finalized by determining the antigenic value of the 331 vaccine candidates using the VaxiJen server to obtain protective antigens prediction.

3.2. Selection of Representative Proteins. Using the calculated characteristics, the number of vaccine candidates was

reduced to 73 proteins, with a stricter selection, as mentioned in the Methods section. These proteins were grouped in seven clusters according to their type and the family to which they belong, as follows: the ESX protein family (3 proteins), the PPE protein family (7 proteins), the PE proteins family (8 proteins), PE_PGRS protein family (16 proteins), lipoproteins (5 proteins), hypothetical proteins (21 proteins), and the final group denominated “others” (13 proteins) (Table 1). This process was carried out because it is well known that members within the same protein family possess a close relationship between their sequences and functions.

With the purpose of selecting representative proteins from each group, we used amino acid sequences from their members to perform alignments using the Clustal X program, with the exception of those from the “others” and hypothetical proteins groups.

In the case of the hypothetical proteins, an alignment was carried out using the Psi-BLAST tool from the NCBI website, in order to grant them a putative function; in some cases, a coincidence was not found, but in others we could assume a probable function (Figure 2).

For the selection of representative proteins, we took into account the similarity among the sequences of the group, the best antigenicity values, the high probability to act as an adhesin, and their conservation in the *M. bovis* and BCG proteome. For this part, we chose 12 representative candidates from the seven protein families (Table 1).

3.3. Immune Response Simulation. The 12 proteins selected were used to conduct simulations of the human immune system response under different conditions with C-ImmSim software. In terms of the results of the simulations, C-ImmSim server showed the following nine graphs for each simulation: B-cell population, B-cell population per state, Th cell population, Th cell population per state, Tc population, Tc population per state, CD population, EP population, and Ab production (Figure 3). We found the same pattern in all the selected proteins with a slight difference among levels. However, we focused mainly on Th1 cells level including all states, because it has been reported that protective immunity against TB is conferred mainly by Th1 cells [14, 29]. We found that the levels of stimulated cells were most similar among the selected proteins when one immunization was performed, but this level improved when the number of immunizations increased, and there was also a remarkable differentiation among the proteins at the final simulation step. PE_PGRS family proteins showed the highest levels of stimulated Th cells, which also generated a good level of B cells stimulation, which is important for complementing the immune response (Figure 4).

3.4. Protein Analysis. The proteins were analyzed individually with several bioinformatics tools to determine whether the protein sequences had a region with important antigenic characteristics, that is, a region where there are matches with hydrophobicity, solvent accessibility, presentation to MHC, and antigenicity.

We did not find a protein that clearly possesses an antigenic region that could be used as an epitope or as a fusion in a vaccine. Conversely, we determined that all the proteins had high values of antigenicity in different parts of the sequence; thus, we recommend the use of complete proteins in a vaccine formulation because using only a fragment could eliminate some epitopes necessary for a complete and protective human immune response against the whole microorganism.

We also found that all of the proteins selected as vaccine candidates could be presented to several MHC with a high probability value, resulting in good probability of immune response induction against these components of *M. tuberculosis*.

In this analysis process, we identified protein subcellular localizations using Phobius tools to confirm the results emitted by NERVE, because Phobius software is more accurate than the program (HMMTOP) utilized by NERVE [21]. This characteristic is important in a vaccine candidate because proteins with cytoplasmic or membrane localization are less antigenic than extracytoplasmic proteins.

3.5. Bibliographic Study. We wanted to know whether the proteins would be safe if we used them on a vaccine formulation prior to the preclinical trials; thus, we studied the information published about different characteristics related with their impact on virulence.

We were able to observe that some proteins have not been studied, but we found information about other members of their family groups, such as PE, PPE, and PE_PGRS proteins, suggesting an influence on immune system evasion and antigenic variation, an important feature in considering a protein that will be included in a vaccine [39, 47, 59]; besides, PE_PGRS family proteins are restricted to pathogenic mycobacteria and, in particular, PE_PGRS11 and PE_PGRS 17 have been reported to induce maturation and activation of human dendritic cells, enhancing the ability of the latter to induce Th cells stimulation [26, 60].

In case of the antigen LppN we did not find specific information, besides, almost all the proteins in *M. tuberculosis* genome lack conserved regions, which means that they are unique proteins with different characteristics. Some lipoproteins are major antigens in the *Mycobacterium* genus that can generate also a cellular and humoral immune response but without immune memory response [50, 61–64].

Erp protein is a virulence factor present only in the *Mycobacterium* genus [51–53], it is an immunodominant antigen related to pathogenicity and is strongly induced in nutrient starvation related to the latency phase [53]. On the other hand, PBPI protein is important in the replication phase because it catalyzes the final steps of bacterial cell wall peptidoglycan synthesis [54, 65, 66].

The EsxL candidate is an ESX-like protein with very similar characteristics to Esat-6, which is an immunodominant secreted protein used in research associated with the diagnosis of TB and new TB vaccines [30]. Esat-6 is a strong T-cell antigen, and its family members are involved in virulence and in host-pathogen interplay via either antigenic variation or antigenic drift [31, 32, 67].

TABLE 1: Proteins selected after reducing parameters and grouped in seven categories according their family group.

Protein family group	ID	Rv	Vaxijen antigenicity value	Length (amino acid)	Psi BLAST
ESX family	gi_15608177_ref_NP_215553.1.	Rv1037c	0.7444	94	ND
	gi_15608338_ref_NP_215714.1.	Rv1198	0.6286	94	ND
	gi_15610755_ref_NP_218136.1.	Rv3619c	0.7444	94	ND
PPE family	gi_57116729_ref_YP_177724.1.	Rv0388c (PPE9)	0.5334	180	ND
	gi_57116916_ref_YP_177840.1.	Rv1801 (PPE29)	0.5636	423	ND
	gi_57116975_ref_YP_177871.1.	Rv2353c (PPE39)	0.897	354	ND
	gi_57117024_ref_YP_177677.1.	Rv2770c (PPE44)	0.5056	382	ND
	gi_57117062_ref_YP_177932.1.	Rv3125c (PPE49)	0.5581	391	ND
	gi_57117064_ref_YP_177934.1.	Rv3135 (PPE50)	0.5029	132	ND
	gi_57117135_ref_YP_177998.1.	Rv3621c (PPE65)	0.5241	413	ND
PE family	gi_57116715_ref_YP_177710.1.	Rv0285 (PE5)	0.6696	102	ND
	gi_57116910_ref_YP_177834.1.	Rv1788 (PE18)	0.6228	99	ND
	gi_57116913_ref_YP_177837.1.	Rv1791 (PE19)	0.6125	94	ND
	gi_57116998_ref_YP_177888.1.	Rv2519 (PE26)	0.718	492	ND
	gi_57117110_ref_YP_177975.1.	Rv3477 (PE31)	0.5325	98	ND
	gi_57117136_ref_YP_177999.1.	Rv3622c (PE32)	0.5099	99	ND
	gi_57117151_ref_YP_178010.1.	Rv3739c (PE67)	0.5101	77	ND
	gi_57117167_ref_YP_178025.1.	Rv3893c (PE36)	0.5971	77	ND
	gi_57116752_ref_YP_177736.1.	Rv0532 (PE_PGRS6)	1.789	594	ND
	gi_57116773_ref_YP_177750.1.	Rv0746 (PE_PGRS9)	1.753	783	ND
	gi_57116787_ref_YP_177759.1.	Rv0832 (PE_PGRS12)	0.6034	137	ND
	gi_57116793_ref_YP_177763.1.	Rv0872 (PE_PGRS15)	2.0866	606	ND
	gi_57116818_ref_YP_177780.1.	Rv1067c (PE_PGRS19)	2.2481	667	ND
gi_57116826_ref_YP_177786.1.	Rv1091 (PE_PGRS22)	2.5016	853	ND	
gi_57116864_ref_YP_177811.1.	Rv1441c (PE_PGRS26)	2.1299	491	ND	
PE_PGRS family	gi_57116905_ref_YP_177831.1.	Rv1795c (wag22)	2.0054	914	ND
	gi_57116924_ref_YP_177847.1.	Rv1840c (PE_PGRS34)	1.5912	515	ND
	gi_57116973_ref_YP_177869.1.	Rv2340c (PE_PGRS39)	1.0512	413	ND
	gi_57117010_ref_YP_177896.1.	Rv2634c (PE_PGRS46)	2.146	778	ND
	gi_57117029_ref_YP_177909.1.	Rv2853 (PE_PGRS48)	2.1046	615	ND
	gi_57117093_ref_YP_177961.1.	Rv3344c (PE_PGRS49)	3.0927	484	ND
	gi_57117098_ref_YP_177965.1.	Rv3367 (PE_PGRS51)	2.0713	588	ND
	gi_57117101_ref_YP_177968.1.	Rv3388 (PE_PGRS52)	2.2486	731	ND
	gi_57117117_ref_YP_177981.1.	Rv3512 (PE_PGRS56)	3.4881	1079	ND

TABLE 1: Continued.

Protein family group	ID	Rv	Vaxijen antigenicity value	Length (amino acid)	Psi BLAST
Lipoproteins	gi_15607723_ref_NP_215097_1	Rv0583c (LpqN)	0.6569	228	ND
	gi_15608368_ref_NP_215744_1	Rv1228 (LpqX)	0.7609	185	ND
	gi_15608679_ref_NP_216057_1	Rv1541c (LprH)	0.5298	197	ND
	gi_15609407_ref_NP_216786_1	Rv2270 (LppN)	0.5899	175	ND
	gi_15609921_ref_NP_217300_1	Rv2784c (LppU)	0.685	171	ND
	gi_15607199_ref_NP_214571_1	Rv0057	0.6907	173	No matches
	gi_157116831_ref_YP_177639_1	Rv1116A	0.6337	91	Related with PE family proteins of <i>Mycobacterium</i>
	gi_15608277_ref_NP_215653_1	Rv1137c	0.76	122	No matches
	gi_15608941_ref_NP_216320_1	Rv1804c	0.575	108	Related with <i>Mycobacterium</i> lipoproteins and gp53 <i>Mycobacterium</i> phage
	gi_15609051_ref_NP_216430_1	Rv1914c	0.5202	135	No matches
	gi_15609215_ref_NP_216594_1	Rv2078	0.5425	104	No matches
	gi_15609220_ref_NP_216599_1	Rv2083	0.8189	314	No matches
Hypotheticals	gi_15609401_ref_NP_216780_1	Rv2264c	0.5862	592	Related with proline and threonine rich <i>Mycobacterium</i> proteins, like chaperone molecular and <i>Rhodococcus</i> hypothetical proteins
	gi_15609420_ref_NP_216799_1	Rv2283	0.7336	64	No matches
	gi_15609429_ref_NP_216808_1	Rv2292c	0.5396	74	No matches
	gi_15609439_ref_NP_216818_1	Rv2302	0.958	80	Related with transduction signal protein and hypothetical proteins from <i>Nocardia</i> and <i>Frankia</i> , DNA binding protein from <i>Streptomyces</i>
	gi_15609797_ref_NP_217176_1	Rv2660c	0.9073	75	No matches
	gi_15609843_ref_NP_217222_1	Rv2706c	0.527	85	No matches
	gi_15610097_ref_NP_217476_1	Rv2960c	0.6945	82	No matches
	gi_15610135_ref_NP_217514_1	Rv2998	0.8339	153	No matches
	gi_15610204_ref_NP_217583_1	Rv3067	0.566	136	No matches
	gi_15610316_ref_NP_217696_1	Rv3180c	0.5182	144	Related with proteins which have pilT domain and with DNA binding proteins from <i>Mycobacterium</i>
	gi_15610343_ref_NP_217723_1	Rv3207c	0.6281	285	Related with membrane proteins and lipoproteins from <i>Corynebacterium</i> and <i>Streptomyces</i>
	gi_157117091_ref_NP_217854_2	Rv3337	0.8824	128	Related with a putative hydrolase from <i>Mycobacterium</i>
gi_15610854_ref_NP_218235_1	Rv3718c	0.7971	147	Related with KanY, polyketide cyclases, and lipid transport proteins from different species	
gi_15611034_ref_NP_218415_1	Rv3898c	0.7675	110	no matches	

TABLE 1: Continued.

Protein family group	ID	Rv	Vaxijen antigenicity value	Length (amino acid)	Psi BLAST
	gi_15607173_ref_NP_214545_1.	Rv0031	0.738	70	ND
	gi_57116685_ref_YP_177687_1.	Rv0050 (PBPI)	0.6113	678	ND
	gi_57116889_ref_NP_216091_2.	Rv1575	0.7415	166	ND
	gi_57117071_ref_YP_177941_1.	Rv3198A	0.6669	84	ND
	gi_15610488_ref_NP_217869_1.	Rv3352c	0.8742	123	ND
	gi_15610946_ref_NP_218327_1.	Rv3810 (Erp)	0.6734	284	ND
	gi_15609895_ref_NP_217274_1.	Rv2758c	0.6574	88	ND
Others	gi_57117060_ref_YP_177930_1.	Rv3118	0.833	100	ND
	gi_15610417_ref_NP_217798_1.	Rv3281	0.7311	177	ND
	gi_15611001_ref_NP_218382_1.	Rv3865	0.604	103	ND
	gi_15609778_ref_NP_217157_1.	Rv2641	0.7934	152	ND
	gi_15607954_ref_NP_215329_1.	Rv0814c	0.833	100	ND
	gi_57116926_ref_YP_177849_1.	Rv1860	0.5244	325	ND

Note: in bold are the highlighted representative proteins selected. ND: not determined.

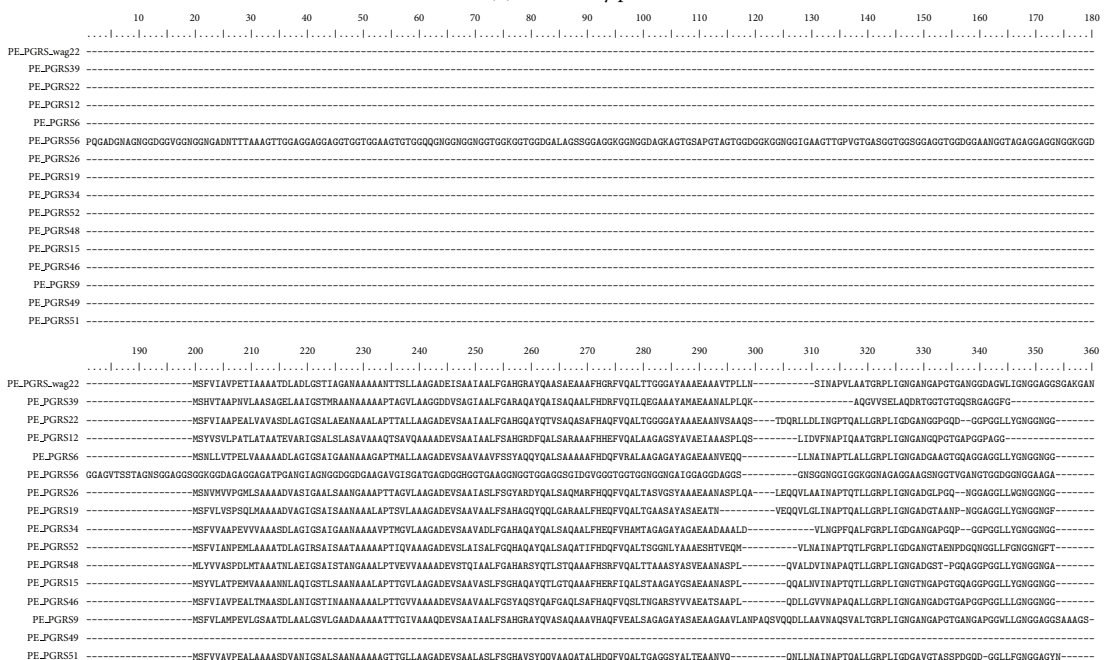
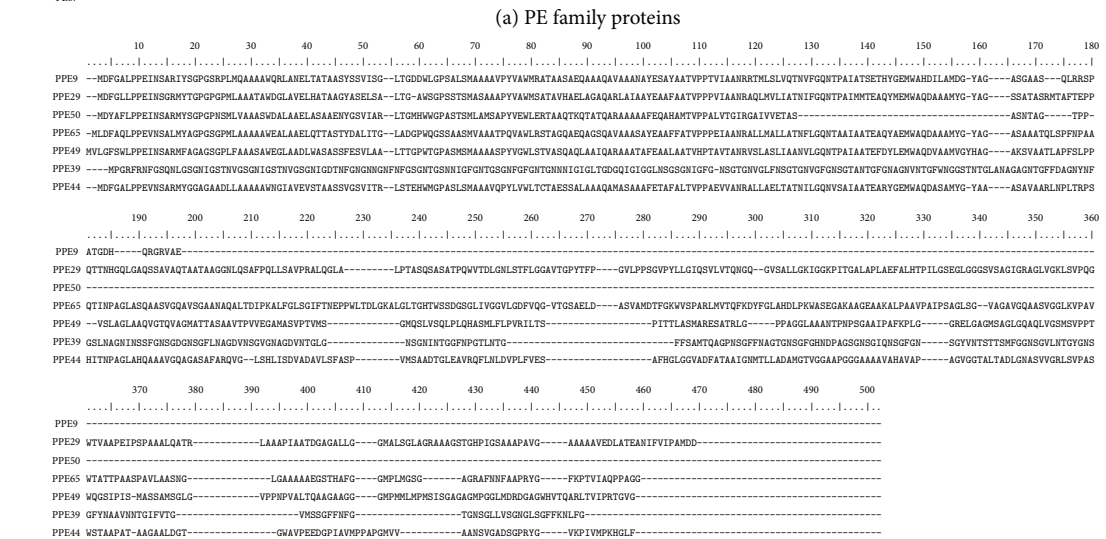
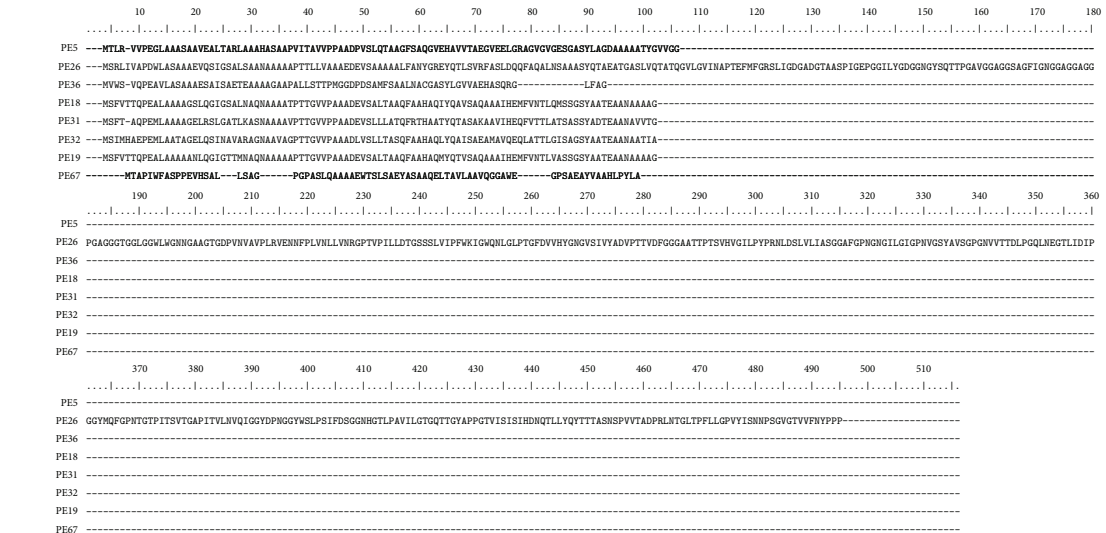


FIGURE 2: Continued.

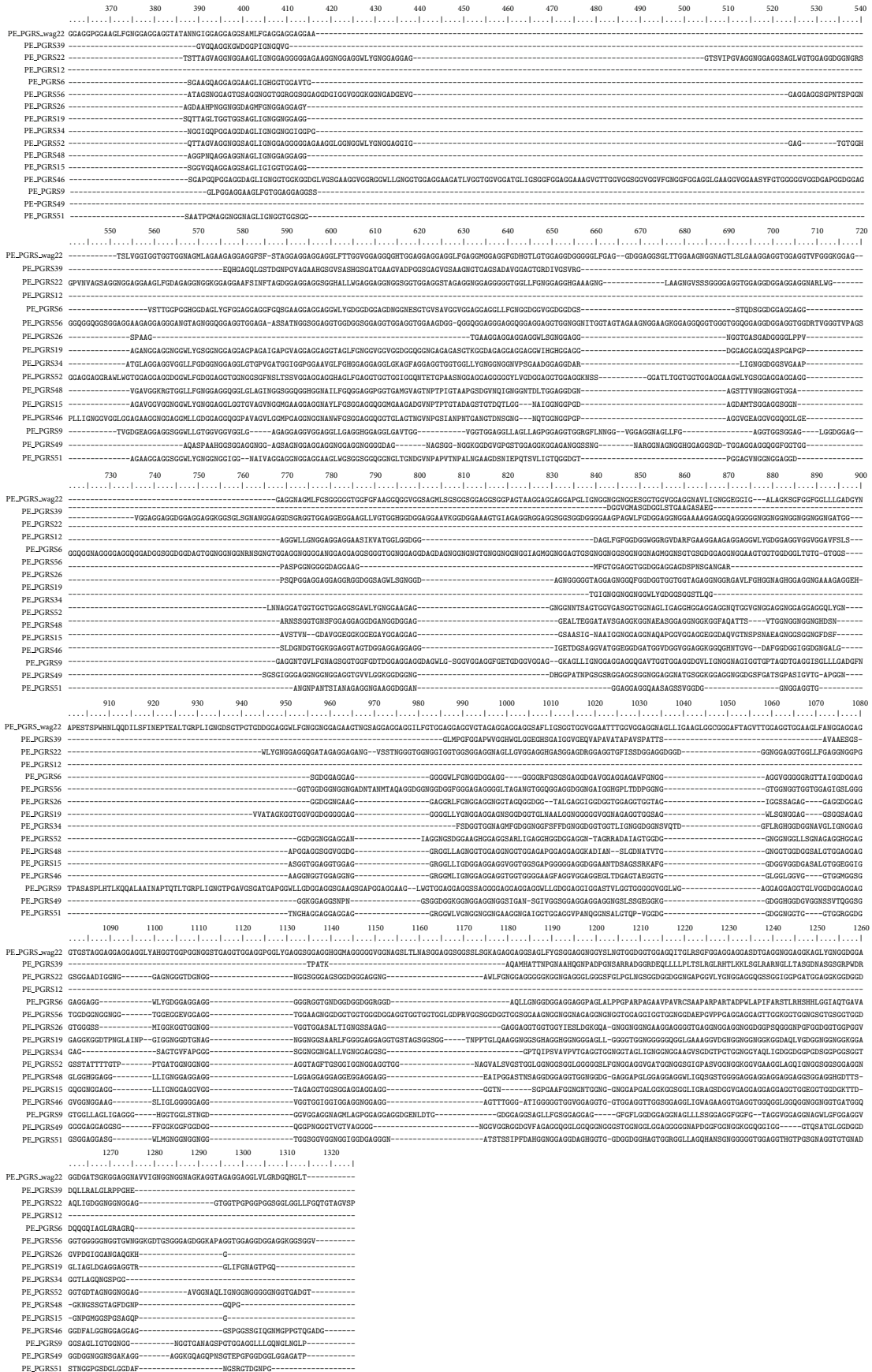


FIGURE 2: Continued.

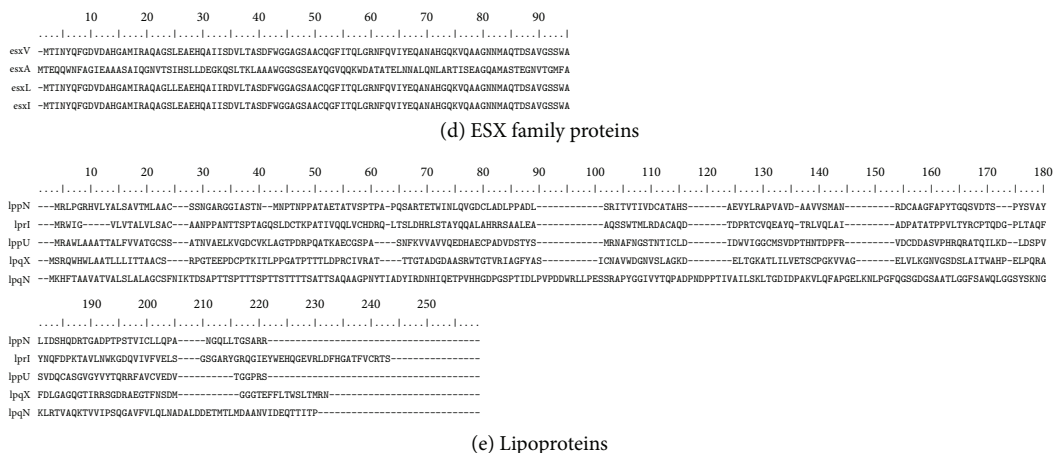


FIGURE 2: Amino acid sequence alignments using Clustal X for the vaccine candidates. The sequences are grouped by protein families and aligned using Clustal X software.

We did not find any information related with hypothetical proteins Rv3207c and Rv3718c, and, up to date, it is unknown if these proteins could be of any risk within a vaccine formulation; thus further studies are needed to elucidate the role of these proteins in this kind of biological products. About the remainder candidates, we found no information that related them with high levels of bacterial virulence; conversely, we identified highlighted features that render these proteins as good vaccine candidates.

3.6. Vaccine Candidate Selection. All parameters studied were instrumental to determine the different and important characteristics to be taken into account for an adequate vaccine candidate. At the end of this study, we selected the following six proteins as the best vaccine candidates: EsxL (NCBI gene locus tag Rv1198), PE26 (NCBI gene locus tag Rv2519), PPE65 (NCBI gene locus tag Rv3621c), PE_PGRS49 (NCBI gene locus tag Rv3344c), PBPI (NCBI gene locus tag Rv0050), and Erp (NCBI gene locus tag Rv3810). These proteins are representative of their family groups and have the best characteristics of antigenic value, adhesin probability, and stimulation of Th cells (Table 2). In addition, the information contained in the publications on these proteins indicates that they could be safe.

These proteins play different roles in mycobacterial pathogenicity and possess featured values of antigenicity and immune response induction; in addition, the PE_PGRS49 candidate is not present in *M. bovis* or in the BCG proteome, an important characteristic that could be useful in the improvement of a specific immune response against pulmonary disease in the current BCG vaccine.

4. Conclusions

The application of bioinformatics programs for the identification of proteins that could be used as vaccine candidates is a very useful, easier, and shorter process compared with traditional vaccinology, which is important for the research concerning public health.

Using RV, we selected six novel vaccine candidates from the *M. tuberculosis* H37Rv proteome, employing mainly *in silico* studies. The six proteins selected: EsxL, PE26, PPE65, PE_PGRS49, PBPI, and Erp correspond to several family proteins and possess different characteristics that are useful and important in vaccine design. The bibliographic information also indicated that they might be safe.

The potential vaccine candidates selected in this work could be used in different vaccine designs to conduct experiments in order to validate them as DNA vaccines, rBCG, or as recombinant proteins, to improve protection against TB, rendering the new vaccines more effective against the pulmonary disease.

We did not propose a specific and unique antigenic region within these protein structures; however, the bioinformatics and bibliographic analyses showed characteristics that make them valuable putative vaccine candidates that could be used further to experimentally investigate whether they are suitable for the development of a new vaccine against tuberculosis.

Conflict of Interests

The authors declare that they have no competing interests.

Authors' Contribution

Gloria P. Monterrubio-López, Jorge A. González-Y-Merchand, and Rosa María Ribas-Aparicio contributed equally to the original concept and design described in the paper. Gloria P. Monterrubio-López performed experiments and contributed in the areas of data collection, analysis, and interpretation. Rosa María Ribas-Aparicio supervised the overall project, providing essential guidance. Gloria P. Monterrubio-López and Rosa María Ribas-Aparicio drafted the paper. All authors revised the paper critically and read and approved the final version.

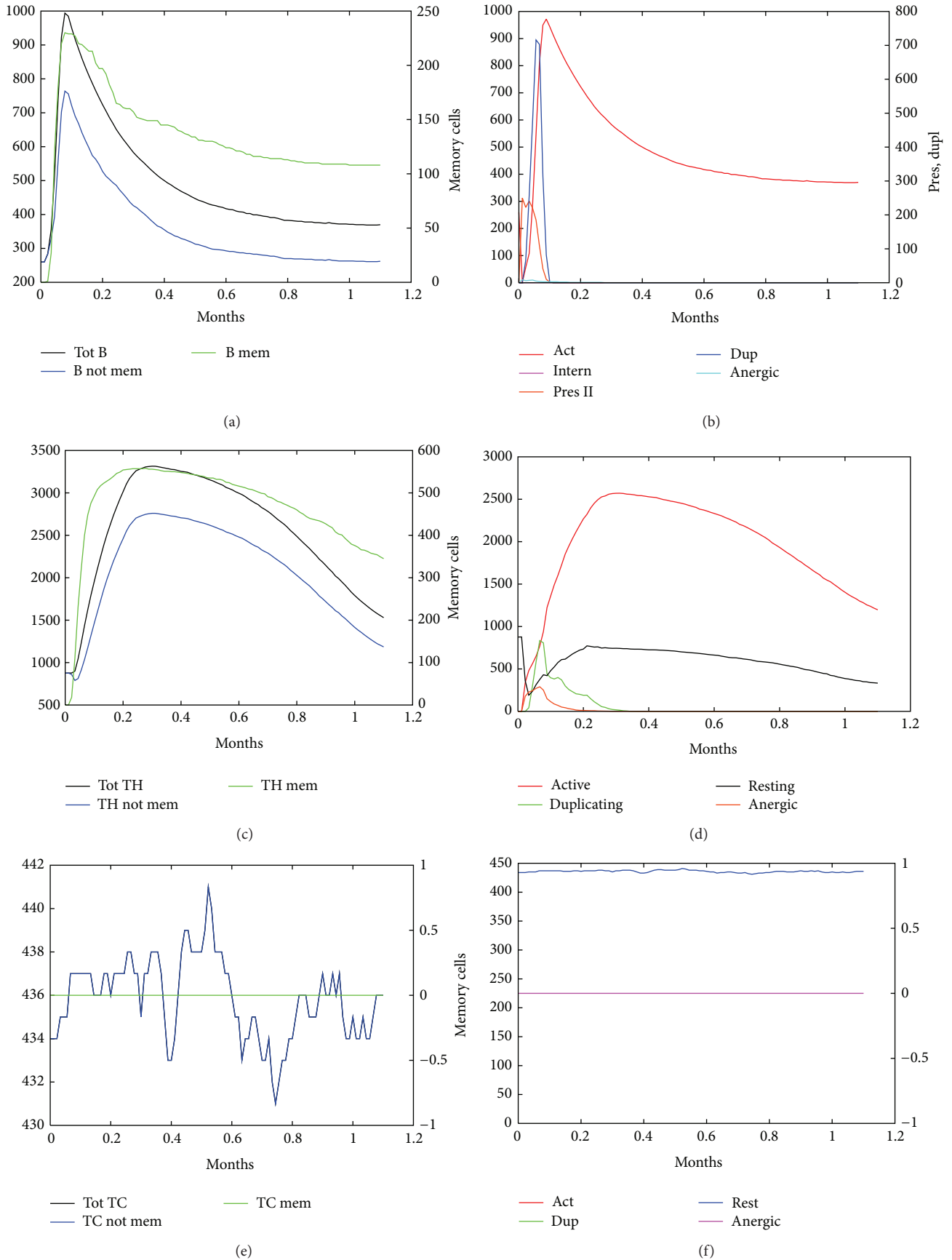


FIGURE 3: Continued.

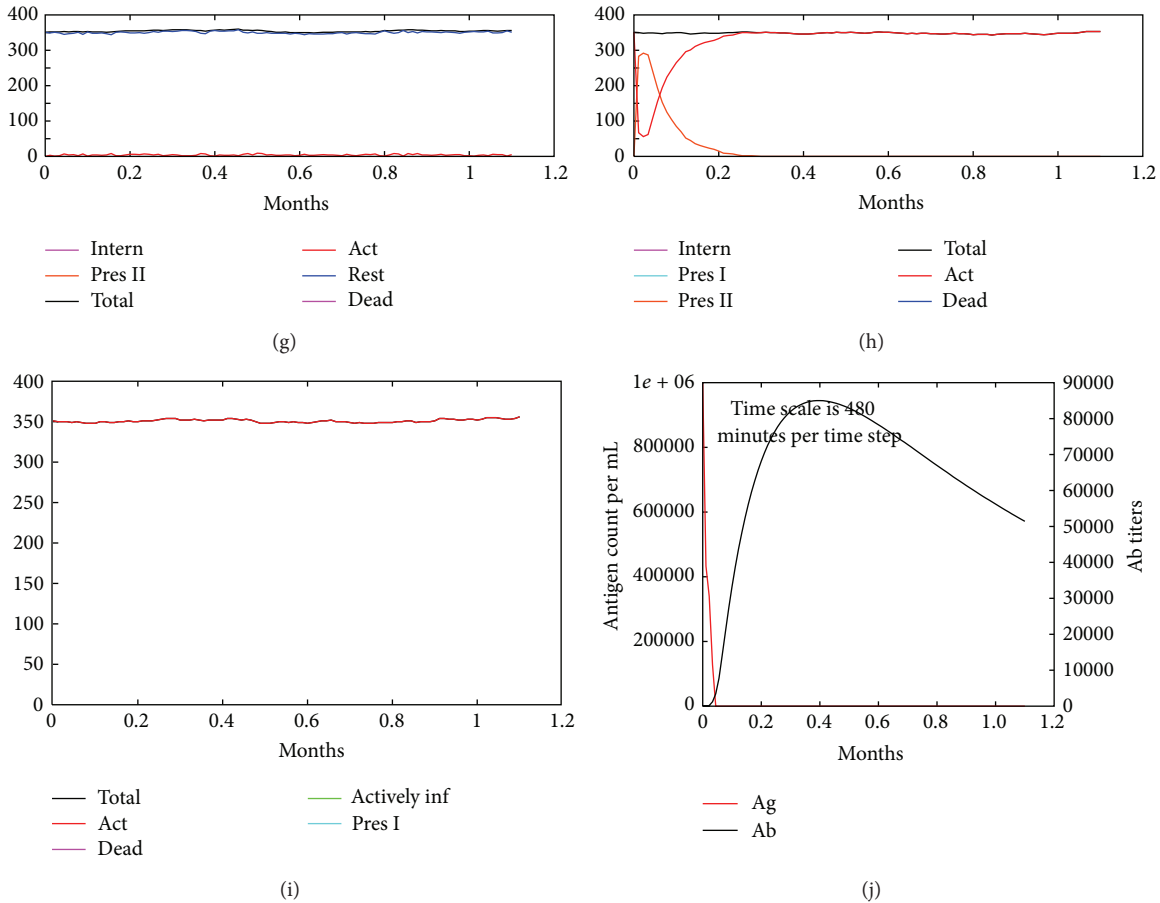


FIGURE 3: C-ImmSim simulation of an immunization experiment using Erp protein. An immunogenic molecule (Erp) was inoculated at time zero. Different cellular populations showed stimulation with Erp antigen at 1.2 months after one dose immunization. (a) B-cell population, (b) B-cell population per state, (c) Th cell population, (d) Th cell population per state, (e) Tc cell population, (f) Tc cell population per state, (g) macrophages population per state, (h) dendritic cell population per state, (i) epithelial cell population, and (j) antibody titers.

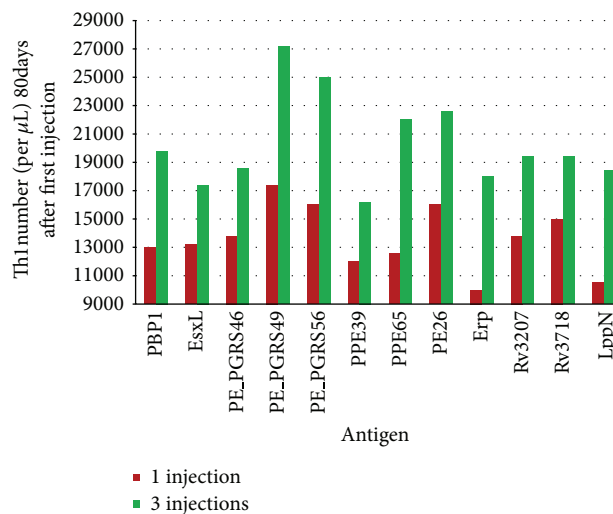


FIGURE 4: Levels of Th cells stimulated with vaccine candidates assessed with the C-ImmSim server. Simulation with C-ImmSim was performed for each vaccine candidate with one and three immunizations, and Th1 cells stimulated values per microliter were identified 80 days after the first immunization. The best stimulation was induced by PE_PGSR49 protein followed by PE_PGSR 56, PE26, PPE65, and PBP1 proteins.

TABLE 2: Vaccine candidates selected by reverse vaccinology.

Characteristic	ESXL	PE_PGERS49	PE26	PPE65	Erp	PBP1
Adhesin probability (%)	80	78	89	82	86	73
Antigenicity value (Vaxijen)	0.6286	3.0927	0.718	0.5241	0.6734	0.6113
Th1 cell number stimulated (1/3 injections)	13200/17400	17400/27200	16000/22600	12600/22000	10000/18000	13000/19800
<i>M. bovis</i> conservation	1.786	0	3.396	6.576	0.271	1.046
BCG conservation	1.786	0	3.298	6.656	0.271	1.046
Immunologic relation	It is suggested to be ESXA similar.	Their protein family members are important antigens recognized by vaccinated and TB patients serum. They are strongly related with antigenic drift and with evasion of immune response.	Their protein family members were used as strain differentiation markers, related with antigenic drift and with evasion of immune response and virulence.		Specific antibodies on cavitary TB patients have been detected.	Not reported
Characteristic or function	It is suggested to be ESXA similar.	The function of their members are varied, some of them are related with granuloma specific expression. They are also related with infectivity influence, necrosis, and apoptosis induction. They are restricted to <i>Mycobacterium</i> pathogenic species.	They are located on 3 conserved PAs and 10% of the <i>Mycobacterium</i> genome is related with this family proteins.		Erp membrane protein precursor, phagosome is related with nutrient reductions and <i>in vivo</i> after infection but before clinical tuberculosis in a rabbit model.	Penicillin binding protein of high molecular weight and membrane bound related with peptidoglycan synthesis. It is expressed only <i>in vivo</i> .
Vaccine trials	Not reported		Not reported		Not reported	Not reported
References	[30–38]	[29, 39–46]	[29, 32, 34, 47–49]		[26, 39, 50–52]	[53–58]

Note: ESXL (ESXL family); PE_PGERS49 (PE_PGERS family); PE26 (PE family); PPE65 (PPE family); Erp and PBP1 (Others).

Acknowledgments

The authors thank Margaret Bruner and Ingrid Mascher for English editing. Rosa María Ribas-Aparicio and Jorge A. González-Y-Merchand are grateful for support of this work from Instituto Politécnico Nacional Grants nos. SIP-20144063 and SIP-20140970. Gloria P. Monterrubio-López thanks CONACyT-México and PIFI-IPN-México for graduate studies fellowships. The authors are also grateful to Dr. Sandro Vivona at Stem CentRx and Dr. Alfonso Méndez-Tenorio, ENCB-IPN professor, for their kind help in installing NERVE and to Dr. Filippo Castiglione at the Istituto per le Applicazioni del Calcolo (IAC) for his valuable advice on the use of the C-ImmSim server.

References

- [1] J. Weiner III and S. H. E. Kaufmann, "Recent advances towards tuberculosis control: vaccines and biomarkers," *Journal of Internal Medicine*, vol. 275, no. 5, pp. 467–480, 2014.
- [2] J. Montoya, J. A. Solon, S. R. C. Cunanan et al., "A randomized, controlled dose-finding Phase II study of the M72/AS01 candidate tuberculosis vaccine in healthy PPD-positive adults," *Journal of Clinical Immunology*, vol. 33, no. 8, pp. 1360–1375, 2013.
- [3] A. P. Junqueira-Kipnis, L. M. Marques Neto, and A. Kipnis, "Role of fused *Mycobacterium tuberculosis* immunogens and adjuvants in modern tuberculosis vaccines," *Frontiers in Immunology*, vol. 5, article 188, 2014.
- [4] World Health Organization, *Global Tuberculosis Report*, World Health Organization, Lyon, France, 2013.
- [5] B. M. Buddle, N. A. Parlane, D. N. Wedlock, and A. Heiser, "Overview of vaccination trials for control of tuberculosis in cattle, wildlife and humans," *Transboundary and Emerging Diseases*, vol. 60, no. 1, pp. 136–146, 2013.
- [6] M. Lu, Z. Y. Xia, and L. Bao, "Enhancement of antimycobacterial Th1-cell responses by a *Mycobacterium bovis* BCG prime-protein boost vaccination strategy," *Cellular Immunology*, vol. 285, no. 1-2, pp. 111–117, 2013.
- [7] S. H. E. Kaufmann, "Tuberculosis vaccine development at a divide," *Current Opinion in Pulmonary Medicine*, vol. 20, no. 3, pp. 294–300, 2014.
- [8] J. Adu-Bobie, B. Capocchi, D. Serruto, R. Rappuoli, and M. Pizza, "Two years into reverse vaccinology," *Vaccine*, vol. 21, no. 7-8, pp. 605–610, 2003.
- [9] R. Ferreira and G. A. Porco, "Vacunas derivadas del análisis de los genomas: vacunología inversa," *Interiencia*, vol. 33, pp. 353–358, 2008.
- [10] M. Scarselli, M. M. Giuliani, J. Adu-Bobie, M. Pizza, and R. Rappuoli, "The impact of genomics on vaccine design," *Trends in Biotechnology*, vol. 23, no. 2, pp. 84–91, 2005.
- [11] C. D. Rinaudo, J. L. Telford, R. Rappuoli, and K. L. Seib, "Vaccinology in the genome era," *The Journal of Clinical Investigation*, vol. 119, no. 9, pp. 2515–2525, 2009.
- [12] A. S. Mustafa, "In silico analysis and experimental validation of *Mycobacterium tuberculosis*-specific proteins and peptides of *Mycobacterium tuberculosis* for immunological diagnosis and vaccine development," *Medical Principles and Practice*, vol. 22, no. 1, pp. 43–51, 2013.
- [13] S. Vivona, F. Bernante, and F. Filippini, "NERVE: new enhanced reverse vaccinology environment," *BMC Biotechnology*, vol. 6, article 35, 2006.
- [14] S. Bertholet, S. G. Reed, and R. Rappuoli, "Reverse vaccinology applied to TB," in *The Art & Science of Tuberculosis Vaccine Development*, N. M. Nor, A. Acosta, and M. E. Sarmiento, Eds., pp. 413–431, Oxford University Press, Pulau Pinang, Malaysia, 2nd edition, 2014.
- [15] N. Rapin, O. Lund, M. Bernaschi, and F. Castiglione, "Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system," *PLoS ONE*, vol. 5, no. 4, Article ID e9862, 2010.
- [16] R. Chaudhuri, D. Kulshreshtha, M. V. Raghunandan, and S. Ramachandran, "Integrative immunoinformatics for Mycobacterial diseases in R platform," *Systems and Synthetic Biology*, vol. 8, no. 1, pp. 27–39, 2014.
- [17] Y. He and Z. Xiang, "Databases and in silico tools for vaccine design," *Methods in Molecular Biology*, vol. 993, pp. 115–127, 2013.
- [18] National Center for Biotechnology Information (NCBI), <http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>.
- [19] I. A. Doytchinova and D. R. Flower, "VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines," *BMC Bioinformatics*, vol. 8, article 4, 2007.
- [20] J. D. Thompson, T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins, "The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools," *Nucleic Acids Research*, vol. 25, no. 24, pp. 4876–4882, 1997.
- [21] L. Käll, A. Krogh, and E. L. L. Sonnhammer, "A combined transmembrane topology and signal peptide prediction method," *Journal of Molecular Biology*, vol. 338, no. 5, pp. 1027–1036, 2004.
- [22] G. Deléage, C. Combet, C. Blanchet, and C. Geourjon, "ANTHEPROT: an integrated protein sequence analysis software with client/server capabilities," *Computers in Biology and Medicine*, vol. 31, no. 4, pp. 259–267, 2001.
- [23] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch, "ExPASy: the proteomics server for in-depth protein knowledge and analysis," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3784–3788, 2003.
- [24] R. Vita, L. Zarebski, J. A. Greenbaum et al., "The immune epitope database 2.0," *Nucleic Acids Research*, vol. 38, no. 1, Article ID gkp1004, pp. D854–D862, 2009.
- [25] H.-G. Rammensee, J. Bachmann, N. P. N. Emmerich, O. A. Bachor, and S. Stevanović, "SYFPEITHI: database for MHC ligands and peptide motifs," *Immunogenetics*, vol. 50, no. 3-4, pp. 213–219, 1999.
- [26] H. Singh and G. P. S. Raghava, "ProPred: prediction of HLA-DR binding sites," *Bioinformatics*, vol. 17, no. 12, pp. 1236–1237, 2002.
- [27] D. R. Flower, *Bioinformatics for Vaccinology*, Wiley-Blackwell, New York, NY, USA, 1st edition, 2008.
- [28] Vaccine Investigation and Online Information Network (VIO-LIN), <http://www.violinet.org/>.
- [29] J. A. Tricca and J. K. Nambiar, "Challenge of developing new tuberculosis vaccines to generate life-long protective immunity," *Expert Review of Vaccines*, vol. 8, no. 7, pp. 823–825, 2009.
- [30] A. Bekmurzayeva, M. Sypabekova, and D. Kanayeva, "Tuberculosis diagnosis using immunodominant, secreted antigens of *Mycobacterium tuberculosis*," *Tuberculosis*, vol. 93, no. 4, pp. 381–388, 2013.

- [31] P. Brodin, I. Rosenkrands, P. Andersen, S. T. Cole, and R. Brosch, "ESAT-6 proteins: protective antigens and virulence factors?" *Trends in Microbiology*, vol. 12, no. 11, pp. 500–508, 2004.
- [32] S. Bertholet, G. C. Ireton, D. J. Ordway et al., "A defined tuberculosis vaccine candidate boosts BCG and protects against multidrug-resistant *Mycobacterium tuberculosis*," *Science Translational Medicine*, vol. 2, no. 53, Article ID 53ra74, 2010.
- [33] D. Bottai and R. Brosch, "Mycobacterial PE, PPE and ESX clusters: novel insights into the secretion of these most unusual protein families," *Molecular Microbiology*, vol. 73, no. 3, pp. 325–328, 2009.
- [34] S. C. Derrick and S. L. Morris, "The ESAT6 protein of *Mycobacterium tuberculosis* induces apoptosis of macrophages by activating caspase expression," *Cellular Microbiology*, vol. 9, no. 6, pp. 1547–1555, 2007.
- [35] N. C. G. van Pittius, S. L. Sampson, H. Lee, Y. Kim, P. D. van Helden, and R. M. Warren, "Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (*esx*) gene cluster regions," *BMC Evolutionary Biology*, vol. 6, article 95, 2006.
- [36] M. Harboe, T. Oettinger, H. G. Wiker, I. Rosenkrands, and P. Andersen, "Evidence for occurrence of the ESAT-6 protein in *Mycobacterium tuberculosis* and virulent *Mycobacterium bovis* and for its absence in *Mycobacterium bovis* BCG," *Infection and Immunity*, vol. 64, no. 1, pp. 16–22, 1996.
- [37] F.-A. Mir, S. H. E. Kaufmann, and A. N. Eddine, "A multicistronic DNA vaccine induces significant protection against tuberculosis in mice and offers flexibility in the expressed antigen repertoire," *Clinical and Vaccine Immunology*, vol. 16, no. 10, pp. 1467–1475, 2009.
- [38] P. S. Renshaw, K. L. Lightbody, V. Veverka et al., "Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6," *The EMBO Journal*, vol. 24, no. 14, pp. 2491–2498, 2005.
- [39] X.-Y. He, J. Li, J. Hao et al., "Assessment of five antigens from *Mycobacterium tuberculosis* for serodiagnosis of tuberculosis," *Clinical and Vaccine Immunology*, vol. 18, no. 4, pp. 565–570, 2011.
- [40] B. Samten, X. Wang, and P. F. Barnes, "*Mycobacterium tuberculosis* ESX-1 system-secreted protein ESAT-6 but not CFP10 inhibits human T-cell immune responses," *Tuberculosis*, vol. 89, no. 1, pp. S74–S76, 2009.
- [41] S. Banu, N. Honoré, B. Saint-Joanis, D. Philpott, M.-C. Prévost, and S. T. Cole, "Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens?" *Molecular Microbiology*, vol. 44, no. 1, pp. 9–19, 2002.
- [42] M. J. Brennan, G. Delogu, Y. Chen et al., "Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells," *Infection and Immunity*, vol. 69, no. 12, pp. 7326–7333, 2001.
- [43] P. P. Singh, M. Parra, N. Cadieux, and M. J. Brennan, "A comparative study of host response to three *Mycobacterium tuberculosis* PE_PGRS proteins," *Microbiology*, vol. 154, no. 11, pp. 3469–3479, 2008.
- [44] M. G. Chaitra, M. S. Shaila, and R. Nayak, "Evaluation of T-cell responses to peptides with MHC class I-binding motifs derived from PE_PGRS 33 protein of *Mycobacterium tuberculosis*," *Journal of Medical Microbiology*, vol. 56, no. 4, pp. 466–474, 2007.
- [45] G. Delogu, C. Pusceddu, A. Bua, G. Fadda, M. J. Brennan, and S. Zanetti, "Rv1818c-encoded PE_PGRS protein of *Mycobacterium tuberculosis* is surface exposed and influences bacterial cell structure," *Molecular Microbiology*, vol. 52, no. 3, pp. 725–733, 2004.
- [46] C. Espitia, J. P. Lacleste, M. Mondragón-Palomino et al., "The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins?" *Microbiology*, vol. 145, no. 12, pp. 3487–3495, 1999.
- [47] G. Delogu and M. J. Brennan, "Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*," *Infection and Immunity*, vol. 69, no. 9, pp. 5606–5611, 2001.
- [48] A. Karboul, N. C. G. van Pittius, A. Namouchi et al., "Insights into the evolutionary history of tubercle bacilli as disclosed by genetic rearrangements within a PE_PGRS duplicated gene pair," *BMC Evolutionary Biology*, vol. 6, article 107, 2006.
- [49] A. S. Mustafa, R. Al-Attayah, S. N. M. Hanif, and F. A. Shaban, "Efficient testing of large pools of *Mycobacterium tuberculosis* RD1 peptides and identification of major antigens and immunodominant peptides recognized by human Th1 cells," *Clinical and Vaccine Immunology*, vol. 15, no. 6, pp. 916–924, 2008.
- [50] R. Al-Attayah and A. S. Mustafa, "Computer-assisted prediction of HLA-DR binding and experimental analysis for human promiscuous Th1-Cell peptides in the 24 kDa secreted lipoprotein (LppX) of *Mycobacterium tuberculosis*," *Scandinavian Journal of Immunology*, vol. 59, no. 1, pp. 16–24, 2004.
- [51] D. Kocincová, B. Sondén, L. De Mendonça-Lima, B. Gicquel, and J.-M. Reyat, "The Erp protein is anchored at the surface by a carboxy-terminal hydrophobic domain and is important for cell-wall structure in *Mycobacterium smegmatis*," *FEMS Microbiology Letters*, vol. 231, no. 2, pp. 191–196, 2004.
- [52] L. I. Klepp, M. Soria, F. C. Blanco et al., "Identification of two proteins that interact with the Erp virulence factor from *Mycobacterium tuberculosis* by using the bacterial two-hybrid system," *BMC Molecular Biology*, vol. 10, article 3, 2009.
- [53] L. de Mendonça-Lima, Y. Bordat, E. Pivert et al., "The allele encoding the mycobacterial Erp protein affects lung disease in mice," *Cellular Microbiology*, vol. 5, no. 1, pp. 65–73, 2003.
- [54] S. Bhakta and J. Basu, "Overexpression, purification and biochemical characterization of a class a high-molecular-mass penicillin-binding protein (PBP), PBP1* and its soluble derivative from *Mycobacterium tuberculosis*," *Biochemical Journal*, vol. 361, no. 3, pp. 635–639, 2002.
- [55] F.-X. Berthet, M. Lagranderie, P. Gounon et al., "Attenuation of virulence by disruption of the *Mycobacterium tuberculosis* *erp* gene," *Science*, vol. 282, no. 5389, pp. 759–762, 1998.
- [56] H. Billman-Jacobe, R. E. Haites, and R. L. Coppel, "Characterization of a *Mycobacterium smegmatis* mutant lacking penicillin binding protein 1," *Antimicrobial Agents and Chemotherapy*, vol. 43, no. 12, pp. 3011–3013, 1999.
- [57] L. G. Dover, A. M. Cerdeño-Tárraga, M. J. Pallen, J. Parkhill, and G. S. Besra, "Comparative cell wall core biosynthesis in the mycolated pathogens, *Mycobacterium tuberculosis* and *Corynebacterium diphtheriae*," *FEMS Microbiology Reviews*, vol. 28, no. 2, pp. 225–250, 2004.
- [58] J. E. Graham and J. E. Clark-Curtiss, "Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 20, pp. 11554–11559, 1999.

- [59] C. R. McEvoy, P. D. Van Helden, R. M. Warren, and N. C. G. Van Pittius, "Evidence for a rapid rate of molecular evolution at the hypervariable and immunogenic *Mycobacterium tuberculosis* PPE38 gene region," *BMC Evolutionary Biology*, vol. 9, no. 1, article no. 237, 2009.
- [60] K. Bansal, S. R. Elluru, Y. Narayana et al., "PE-PGRS antigens of *Mycobacterium tuberculosis* induce maturation and activation of human dendritic cells," *The Journal of Immunology*, vol. 184, no. 7, pp. 3495–3504, 2010.
- [61] M. G. Drage, N. D. Pecora, A. G. Hise et al., "TLR2 and its co-receptors determine responses of macrophages and dendritic cells to lipoproteins of *Mycobacterium tuberculosis*," *Cellular Immunology*, vol. 258, no. 1, pp. 29–37, 2009.
- [62] A.-H. Hovav, J. Mullerad, L. Davidovitch et al., "The *Mycobacterium tuberculosis* recombinant 27-kilodalton lipoprotein induces a strong Th1-type immune response deleterious to protection," *Infection and Immunity*, vol. 71, no. 6, pp. 3146–3154, 2003.
- [63] I. C. Sutcliffe and D. J. Harrington, "Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components," *FEMS Microbiology Reviews*, vol. 28, no. 5, pp. 645–659, 2004.
- [64] M. Rezwan, T. Grau, A. Tschumi, and P. Sander, "Lipoprotein synthesis in mycobacteria," *Microbiology*, vol. 153, no. 3, pp. 652–658, 2007.
- [65] E. C. Hett and E. J. Rubin, "Bacterial growth and cell division: a mycobacterial perspective," *Microbiology and Molecular Biology Reviews*, vol. 72, no. 1, pp. 126–156, 2008.
- [66] E. C. Hett, M. C. Chao, A. J. Steyn, S. M. Fortune, L. L. Deng, and E. J. Rubin, "A partner for the resuscitation-promoting factors of *Mycobacterium tuberculosis*," *Molecular Microbiology*, vol. 66, no. 3, pp. 658–668, 2007.
- [67] P. Brodin, L. Majlessi, L. Marsollier et al., "Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence," *Infection and Immunity*, vol. 74, no. 1, pp. 88–98, 2006.