Tuberculosis 94 (2014) 475–481

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: http://intl.elsevierhealth.com/journals/tube

IMMUNOLOGICAL ASPECTS

Immunoinformatics study on highly expressed *Mycobacterium tuberculosis* genes during infection

Le Thuy Nguyen Thi [a], Maria Elena Sarmiento [b], Romel Calero [c], Frank Camacho [d], Fatima Reyes [d], Md Murad Hossain [b], Gustavo Sierra Gonzalez [c], Mohd Nor Norazmi [b, e], Armando Acosta [e, *]

[a] *Biotechnology Centre of Ho Chi Minh City, 176 Hai Ba Trung Street, Ho Chi Minh City, Viet Nam*
[b] *PPSK, Health Campus, Universiti Sains Malaysia, Malaysia*
[c] *Finlay Institute, Cuba*
[d] *Universidad de Concepcion, Chile*
[e] *INFORMM, Health Campus, Universiti Sains Malaysia, Malaysia*

ARTICLE INFO

Article history:
Received 16 January 2014
Received in revised form
4 June 2014
Accepted 8 June 2014

Keywords:
Mycobacterium tuberculosis
Gene expression
Epitope prediction

SUMMARY

The most important targets for vaccine development are the proteins that are highly expressed by the microorganisms during infection *in-vivo*. A number of *Mycobacterium tuberculosis* (Mtb) proteins are also reported to be expressed *in-vivo* at different phases of infection. In the present study, we analyzed multiple published databases of gene expression profiles of Mtb *in-vivo* at different phases of infection in animals and humans and selected 38 proteins that are highly expressed in the active, latent and reactivation phases. We predicted T- and B-cell epitopes from the selected proteins using HLAPred for T-cell epitope prediction and BCEPred combined with ABCPred for B-cell epitope prediction. For each selected proteins, regions containing both T- and B-cell epitopes were identified which might be considered as important candidates for vaccine design against tuberculosis.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Tuberculosis (TB) still remains a major health problem worldwide. *Mycobacterium tuberculosis* (Mtb), the causative organism of TB in humans infects 9 million individuals and kills 1.4 million people annually [1,2]. The World Health Organization (WHO) has estimated that one-third of the world's population has already been infected with Mtb [1,2]. The attenuated live *Mycobacterium. bovis* bacille calmette–guerin (BCG) is the vaccine in use for the prevention of TB [3]. However, the variable efficacy of BCG vaccine against TB in adults, and the emergence of multi-drug resistant and extensively drug resistant Mtb strains have made the development of new or improved TB vaccines more urgent [3,4]. The complete genome sequences of Mtb, BCG and other mycobacteria have provided an enormous flow of valuable information that can be useful for designing new vaccine strategies [5–17].

The application of *in silico* techniques in vaccine design is an exciting approach for the discovery of new vaccines as well as for the improvement of existing vaccines [18]. Over the past decade, a number of bioinformatics tools have been developed that use genomic sequence data to predict which parts of a microbe that the immune system will react to, the so-called epitopes, which are the most ideal components of vaccines [19]. After deciphering the complete genome sequence of Mtb and with the use of highly specific and sensitive technologies, such as microarrays and quantitative real-time polymerase chain reaction (RT-PCR), the expression *in-vivo* of genes of Mtb has been studied [17,20–38]. Understanding how Mtb regulates its different gene functions according to environmental changes will probably lead to the understanding of many interesting aspects of the growth patterns of the organism, including activation, latency and reactivation of TB [20–38]. The combination of *in-vivo* and *in silico* studies allows focusing on a smaller group of relevant antigens to be used in vaccine development [18,19]. In this study, we selected Mtb genes that are highly expressed *in-vivo* in order to establish a pre-selected pool of antigens for the search of potential protective epitopes.

* Corresponding author. INFORMM, Health Campus, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia. Tel.: +60 9 767 2402, +60 9 767 2403, +60 9 767 2404; fax: +60 9 764 8673.
E-mail addresses: lenacns83@yahoo.com (L.T. Nguyen Thi), sarmari14@gmail.com (M.E. Sarmiento), rcalero@finlay.edu.cu (R. Calero), frank.camacho@gmail.com (F. Camacho), fatima8526@gmail.com (F. Reyes), mmhossaink140@yahoo.com (M.M. Hossain), gsierra@finlay.edu.cu (G.S. Gonzalez), norazmimn@usm.my (M.N. Norazmi), ducmar13@gmail.com (A. Acosta).

http://dx.doi.org/10.1016/j.tube.2014.06.004
1472-9792/© 2014 Elsevier Ltd. All rights reserved.

2. Materials and methods

2.1. Identification and selection of in-vivo expressed genes of Mtb

A search for the articles reporting *in-vivo* expression of Mtb genes in humans and animals was carried out using the Internet Google Search Tool [20–38]. The genes reported that are highly expressed in humans and animals at all stages of infection: activation, latency and reactivation were then selected for epitope prediction [20–38].

2.2. Prediction of subcellular localization of Mtb proteins

The subcellular localization of the selected Mtb proteins were defined according to the report of the identification and localization of 1044 Mtb proteins using two-dimensional capillary high-performance liquid chromatography coupled with mass spectrometry (2DLC/MS) methods [39]. For the rest of the selected proteins that were not included in the first list of 1044 proteins, we used TBpred server (http://www.imtech.res.in/raghava/tbpred/) for prediction of subcellular localization.

2.3. Epitope prediction

The amino acid sequences and the features of the selected genes of Mtb 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/) and were used for T- and B-cell epitope prediction.

2.3.1. T-cell epitope prediction

T-cell epitopes for the selected Mtb genes were predicted using the HLAPred server (http://www.imtech.res.in/raghava/hlapred/). All the alleles from the list of 51 HLA class I and 36 HLA class II were selected for the prediction of T-cell epitopes.

2.3.2. B-cell epitope prediction

A combination of two servers, Bcepred (http://www.imtech.res.in/raghava/bcepred/) and ABCpred (www.imtech.res.in/raghava/abcpred/), were used for linear B-cell epitope prediction. For Bcepred server: a combination of 7 physico-chemical properties of amino acids (hydrophilicity, flexibility, accessibility, polarity, exposed surface and turns and antigenic propensity) with a threshold of 2.38 was used. For ABCpred server: a threshold of 0.5 and the predicted B-cell epitope length of 16 amino acids were used.

For the selection of the immunogenic regions having both T- and B-cell epitopes, the regions that contain high score predicted B-cell epitope(s) (score $\geq$ 0.7 by ABCpred in accordance to Bcepred prediction result) and are predicted as “Binder” and promiscuous (epitopes presented by multiple HLA alleles) in T-cell epitope prediction (by HLAPred) were trimmed and listed.

3. Results

3.1. Identification and selection of in-vivo highly expressed genes of Mtb

We analyzed 19 published datasets of *in-vivo* expression of Mtb genes during different phases of infection in animals and humans: eight from humans, ten from animals (mainly in mice and rabbits) and one from both humans and mice [20–38]. These studies were directed to identify the genes of Mtb probably responsible for the adaptation and evasion of the host immune response and drug resistance as well as for the reactivation of infection and invasion of the bacteria. The expression profile of the 3924 protein encoding genes from Mtb was analyzed under different conditions during different stages of infection in different hosts [20–38].

From the whole genome of Mtb, 38 genes that were reported to be significantly up-regulated at the active, latent and reactivation phases were selected. We obtained the amino acid sequences and characteristics of the selected genes from the Reference Sequence (RefSeq) database and identified the predicted subcellular localization of the corresponding proteins. Subcellular localization prediction provide useful information about a particular protein and allows to make inferences on the protein's function, to annotate the corresponding genes and genomes, and in particular, in the case of proteins of bacterial pathogens, to identify potential diagnostic, drug and vaccine targets. The list of 38 selected proteins of Mtb categorized by their cellular location, annotation, and function is presented in Table 1.

Among the 38 selected genes, *psts1* (encoding the 38-kDa lipoprotein), *fbpA*, *fbpB* (encoding antigen 85A and antigen 85B, respectively), *esat-6* (encoding a small 6 kDa molecular weight early secreted antigen target 6, Esat-6) and *cfp-10* (encoding for culture filtrate protein 10) are secreted proteins identified to be up-regulated at the early stages of infection. These antigens induce strong immune responses to infection with Mtb or *Mycobacterium bovis*, and they elicit protective immunity in animal models of TB [40]. The high-level of transcription of *esat-6* was also detected in non-replicating tubercle bacilli [25].

Highly expressed genes of Mtb in the latent phase mostly belong to the initial two-component response regulator (DosR) which is activated at the beginning of exposure to hypoxic conditions and the enduring hypoxic response (EHR) system that is activated later during hypoxia [35,41]. DosR regulon is responsible for the transcriptional changes during oxygen limitation, which is considered an important stimulus for the entry of Mtb into a dormant state [41]. Independent of DosR, EHR consists of a larger group of genes significantly induced for longer periods of time in response to hypoxia [35]. Moreover, EHR shows a substantial overlap with Mtb genes induced by nutrient deprivation [35]. The EHR may contain the machinery used to enter into and survive latency [35]. The induced genes during this persistent stage include those that encode for heat shock proteins (*hsp, hspX*), PE and PPE proteins (*Rv0387c*, *PE11*, *PPE54*), cellular metabolism and transport proteins (*icl, narX*, *Rv3406*, *psts1*, *fdxA*), transcriptional regulatory proteins (*Rv1255*, *whiB6*), cellular processes and conserved hypothetical proteins [35].

An important phase of TB is the reactivation of the intracellular tubercle bacilli during immune suppression of the host. Mtb possesses five genes with significant homology to the resuscitation promoting factor (Rpf) of *Micrococcus luteus* [42,43]. Rpfs are small proteins found in many high G + C gram-positive organisms. In picomolar concentration *in-vitro* the Rpf proteins of Mtb have been shown to stimulate the growth of extended-stationary-phase cultures of BCG [42,43]. It is suggested that Rpfs may play a role in dormancy or reactivation of tubercle bacilli *in-vivo*, and could be a potential target for vaccine intervention [44,45]. More recently, Mtb-infected human tissues have been shown to express these proteins [38].

Proteins from the whole genome and the selected proteins of Mtb were classified into 10 functional categories adapted from TubercuList web server (http://genolist.pasteur.fr/TubercuList/). The percentages of the proteins in each category are shown in Figure 1. A large part of the 38 selected proteins are related to conserved hypothetical proteins (38%) and cell wall synthesis and cellular processes (26%). These groups of proteins are over represented in our study compared to the other protein groups encoded by the genome of Mtb.

Download English Version:

https://daneshyari.com/en/article/2401556

Download Persian Version:

https://daneshyari.com/article/2401556

Daneshyari.com