Available online at www.sciencedirect.com
SCIENCE DIRECT
Microbes and Infection 8 (2006) 738–746
Microbes and Infection
www.elsevier.com/locate/micinf

Original article

A combined immuno-informatics and structure-based modeling approach for prediction of T cell epitopes of secretory proteins of *Mycobacterium tuberculosis*

J. Vani [a], M.S. Shaila [a], N.R. Chandra [b], R. Nayak [a,*]

[a] *Department of Microbiology & Cell Biology, Indian Institute of Science, Bangalore 560 012, India*
[b] *Bioinformatics Center, Indian Institute of Science, Bangalore 560 012, India*

Received 2 July 2005; accepted 12 September 2005
Available online 18 January 2006

Abstract

The role of secretory proteins of *Mycobacterium tuberculosis* in pathogenesis and stimulation of specific host responses is well documented. They are also shown to activate different cell types, which subsequently present mycobacterial antigens to T cells. Therefore identification of T cell epitopes from this set of proteins may serve to define candidate antigens with vaccine potential. Fifty-two secretory proteins of *M. tuberculosis* H37Rv were analyzed computationally for the presence of HLA class I binding nonameric peptides. All possible overlapping nonameric peptide sequences from 52 secretory proteins were generated in silico and analyzed for their ability to bind to 33 alleles belonging to A, B and C loci of HLA class I. Fifteen percent of generated peptides are predicted to bind to HLA with halftime of dissociation $T_{1/2} \geq 100$ min and 73% of the peptides predicted to bind are mono-allelic in their binding. The structural basis for recognition of nonamers by different HLA molecules was studied employing structural modeling of HLA class I–peptide complexes and there exists a good correlation between structural analysis and binding prediction. Pathogen peptides that could behave as self- or partially self-peptides in the host were eliminated using a comparative study with the human proteome, thus reducing the number of peptides for analysis. The implications of the finding for vaccine development are discussed vis-à-vis the limitations of the use of subunit vaccine and DNA vaccine.

Keywords: *Mycobacterium tuberculosis*; Secretory proteins; T cell epitopes; HLA class I binding; Vaccine design

1. Introduction

Tuberculosis is a widespread ancient disease, which causes ~3 million deaths with ~8 million new cases emerging each year. The only vaccine that has been used against tuberculosis is BCG, which exhibits wide variations in its efficacy as a prophylactic vaccine against adult pulmonary tuberculosis [1]. Such variations in efficacy have been attributed to several factors such as production methods, bacterial viability, dosage and method of administration, differences in vaccine and geographical location [2–7], genetic differences amongst various population groups [8,9] and interaction with environmental mycobacteria [10]. However, the distribution of HLA polymorphism which differs widely in different populations [11] has not been taken into account. Given the global tuberculosis emergency, it is crucial to elucidate alternative strategies for developing a more promising vaccine useful across diverse populations.

Among the different groups of proteins of *Mycobacterium tuberculosis* that have been focused on as vaccine candidates so far, secretory proteins released into culture filtrate have been studied extensively. Their ability to induce cell-mediated immune responses and protective immunity has been shown both in mouse and guinea pig models of pulmonary tuberculosis [12]. Moreover, culture filtrate proteins of *M. tuberculosis* contain multiple antigens that are strongly recognized by T cells [13]. Recent evidence suggests that the secretory proteins have immunomodulatory activities, and are involved in mycobacterial subversion of the host immune defense [14]. These proteins also play an important role

* Corresponding author. Tel.: +91 80 22932703; fax: +91 80 23602697.
E-mail address: nayak@mcbl.iisc.ernet.in (R. Nayak).

1286-4579/$ - see front matter © 2005 Elsevier SAS. All rights reserved.
doi:10.1016/j.micinf.2005.09.012

in progression of pathogenesis [15,16]. Despite the fact that many secretory proteins such as ESAT-6, Ag 85A, B, C, 19 kDa lipoprotein, and CFP-10 have been shown to be responsible for protective immunity, most other secretory proteins have not been tested systematically.

Evidence suggesting the role of $CD8^+$ T cells in resistance against tuberculosis has been put forth [17] in addition to the known roles of $CD4^+$ T cells and cytokines in limiting mycobacterial infections. Further, MHC class I restricted $CD8^+$ T lymphocytes specific for mycobacterial antigens have been observed in mouse models of TB [18] as well as in humans [19–21] thus emphasizing the role of $CD8^+$ T cell-mediated immune responses. Since $CD8^+$ T cells recognize peptides presented through MHC class I [22–24] it becomes important to identify all class I binding peptides from candidate T cell antigens such that a number of HLA alleles are included to ensure the total population coverage. In theory, identification of HLA-binding peptides is the first step in the efforts to identify all T cell antigens of *M. tuberculosis* and to develop an effective vaccine against tuberculosis.

Relatively very few $CD8^+$ restricted epitopes have been identified against human MHC alleles in *M. tuberculosis* [25]. Since the genome sequence of *M. tuberculosis* is available [26], it is possible to identify proteins and peptides with vaccine potential from the proteome that could provide immune protection in individuals with different HLA backgrounds.

In the present work, with the aim of identifying candidate antigens for vaccine design, we have carried out an in silico analysis of those proteins, which are likely to be secreted by *M. tuberculosis* into the extracellular environment. Fifty-two putative secretory proteins of *M. tuberculosis*, which have been defined as secretory based on the presence of amino terminal signal sequences [27], were chosen and a set of HLA class I binding peptides were identified. We also tested the binding of nonamers of mycobacterial proteins with reported CTL activity using the same algorithm, as a validation step. Further, molecular modeling and structural analysis were used as an additional approach to ascertain the structural feasibilities and interaction of predicted peptides with MHC.

2. Methods

2.1. Prediction of MHC class I binding peptides

Sequences of 52 putative proteins that have been identified as secretory proteins, based on the presence of amino terminal signal sequence and devoid of membrane-anchoring moieties and are secreted via general export pathway [27], were obtained from Sanger center database (http://www.sanger.ac.uk/projects/*M.tuberculosis*). Although a comprehensive identification of secretory proteins in *M. tuberculosis* is not available in the literature as yet, this list is expected to form a significant portion of possible secretory proteins and therefore facilitates a first effort of systematic screening of HLA-binding peptides from secretory proteins. All possible overlapping nonamers were generated systematically from each of these sequences and were analyzed for their potential to bind to 33 different alleles of HLA-I molecules, using a prediction algorithm from BIMAS (http://bimas.dcrt.nih.gov/molbio/hla_bind), which identifies and ranks 8-mer, 9-mer, and 10-mer peptides that contain allele specific binding motifs for HLA class I alleles measured in terms of halftime of dissociation $T_{1/2}$ of β_2 microglobulin [28]. This method, which is based on regression analysis and trained with experimentally derived peptide binding matrices, gives quantitative predictions in terms of half-lives for the dissociation of β_2 microglobulin from the MHC complex. Based on the available experimental data, this prediction algorithm estimates binding against 33 HLA class I alleles, which include 9 HLA-A alleles, 20 HLA-B alleles and 4 HLA-C alleles. Since the 9aa length peptides are considered to be optimum for HLA class I binding [22–24], the algorithm was set to generate all possible overlapping nonamers. All the nonamers in each protein were scanned against each of the 33 HLA alleles available for prediction using BIMAS. The binding was estimated in terms of halftime of β_2 microglobulin dissociation rate (at default cutoff, $T_{1/2}$ of $\geq$100 min). Only those peptides, which were predicted to bind to any of the 33 alleles, were chosen for further analysis. Though the cutoff of 100 min was chosen arbitrarily, it is hoped that it is conservative enough to provide a set of high-affinity peptides derived from diverse proteins, and when included in the vaccine cocktail would make sure of the recognition of those antigens which are synthesized in lesser amounts by the bacteria.

As a validation exercise, the prediction was tested by comparing the scores of the experimentally mapped CTL epitopes from known T cell response inducing antigens of *M. tuberculosis*, against specific HLA alleles using BIMAS algorithm. Another validation exercise was to either deliberately change the peptide sequences or input overlapping peptide fragment that contained only part of the good binder sequence. The algorithm correctly predicted that the modified peptide either bound very poorly or did not bind at all, thus providing good negative controls.

2.2. Identification of ''self-'' peptides

Peptides predicted to bind to HLA were analyzed for homology with any of the human proteins annotated so far, using the BLAST algorithm. Each of the 9-mer peptides was checked with peptides derived from each of the 45,513 human ORFs [29,30]. In view of the short length of peptides, filtering for the low complexity regions was removed, while increasing the *e*-value cutoff. The BLAST results were then passed with a perl script (locally developed) to identify those peptides that exhibited (i) 100% (9aa), (ii) 90% (8aa), (iii) 70% (7aa), and (iv) less than 70% identities with the nonamer peptides derived from the human proteome.

2.3. Molecular modeling of peptide–HLA complexes

The structural feasibility of replacement of the resident peptide with predicted HLA-binding peptides from *M. tuberculosis* was analyzed employing the three dimensional crystal structures of HLA–peptide complex as templates. Crystal structures of