In silico analysis of subcellular localization of putative proteins of Mycobacterium tuberculosis H37Rv strain

P Somvanshi, V Singh, P Seth

Abstract

Mycobacterium tuberculosis is a facultative intracellular pathogen that has evolved the ability to survive and multiply within human macrophages. These bacteria comprise of significant proteins, which were involved in the pathogenesis and regulation of cell activity. The insilico prediction of protein subcellular localization has been extensively studied for prokaryotic, virus and eukaryotic organisms. But, in the case of Mycobacterium, proteins are often involved in extensive interactions at various subcellular localizations in cell. Total thirty-nine putative proteins of M. tuberculosis were predicted for four locations viz cytoplasmic, integral membrane, secretory and protein attached to membrane by Lipid anchor in the subcellular localization. Such predictions provide a method to annotate Mycobacterium proteomes with subcellular localization information rapidly. They have widespread applications in function of proteins in the host cell and in designing the tuberculosis drugs.

INTRODUCTION

Tuberculosis is a global problem and its suffering ranges from less than 10 per 100,000 in North America, 100-300 per 100,000 in Asia and Western Russia to over 300 per 100,000 in Southern and Central Africa. In every 15 seconds there is one death from tuberculosis (2 million deaths per year) and 8 million people develop tuberculosis annually, without treatment up to 60% people infected will be dying. Its major rationales were poverty, lack of healthy living conditions and adequate medical care (Smith, 2003). Tuberculosis continues to affect about 30% world’s population, predominantly in developing countries, despite existence of chemotherapeutic drugs and widespread use of the Mycobacterium bovis BCG vaccine. Effective chemotherapeutic treatment takes long time, it is expensive, and not available to people in various parts of world where needed most. The situation is further complicated by appearance of multidrug-resistant strains. BCG vaccination efficacy is also controversial, as it is not succeeded to protect adults against pulmonary tuberculosis (Bloom and Murray 1992; Bloom and McKinney 1999).

For predicting sub cellular location of eukaryotic, prokaryotic (Gram-negative and Gram-positive bacteria) various methods had been developed but no method has been develop for mycobacterium protein, which may represent repertoire of potent immunogens of this dreaded pathogen. In this analysis, attempts were made to develop a method for prediction of sub cellular location of Mycobacterium proteins. This group of organism is well known for its pathogenicity. After BCG developed in 1921, till date we do not have any promising vaccine against tuberculosis. Furthermore, several new pharmaceutical targets unravel to combat against the multi-drug resistant strains of Mycobacterium. One of the key features of Gene ontology is cellular localization, which gives important information about a protein (Lodish et al 1995).

Earlier, cellular localization of M. tuberculosis is based on in vitro assay like ELISA, western blotting and in situ. In 52 proteins of TM1 subgroup-only 7 had been previously reported to be secretary proteins (Gomez et al 2000). Seven novel antigens of M. tuberculosis, previously identified based on its reactivity to sera from patients with tuberculosis, were characterized. One protein identity was localized in membranes and two were cytosolic, while two others, which had a high proline contents, were tightly associated with the cell wall one protein was secreted (Amara et al 1998).

Thus, it is important to predict subcellular localization of protein in pathogenic organism like Mycobacterium.
Generally, existing methods of subcellular localization developed for eukaryotic proteins like TSSub, LOCSVMPSI, ESLpred, Euk-Ploc (Guo and Lin 2006; Xie et al 2005; Bhasin and Raghava 2004; Shen et al 2007). As various bioinformatics tools were available for prediction of subcellular localization of prokaryotic proteins viz. PSORTb, PSLpred (Gardy et al 2005; Bhasin et al 2005). A model has been developed for predicting four subcellular locations of mycobacterium proteins, namely cytoplasmic, integral membrane, secretory and membrane-attached proteins (Bendtsen et al 2004). No bioinformatics tool was present for predicting subcellular localization of Mycobacterial proteins until now. Recently, TBpred have been developed for the prediction of Mycobacterium protein subcellular localization. In the present study, this online server for prediction of four subcellular locations like cytoplasmic, secretory and protein attached to membrane by lipid anchor and integral membrane of putative proteins were used. The aim behind this study was to predict the subcellular localization of putative proteins of M. tuberculosis H37Rv strain as they might be useful for targeting antimycobacterial drugs.

MATERIALS AND METHODS

COLLECTION OF SEQUENCES

The complete nucleotide and protein sequences of culture filtrate, cell surface, lipid & fat metabolism, amino acid & purine biosynthesis genes, anaerobic respiration & oxidative stress, metal uptake of Mycobacterium tuberculosis H37Rv were extracted from biological database National Centre for Biotechnology Information (NCBI) cited at http://www.ncbi.nlm.nih.gov

ANALYSIS OF PHYSICO CHEMICAL PROPERTIES

The physico-chemical properties of proteins were analyzed viz. total number of amino acids, molecular weight and isoelectric point with Generunner, DNAstar and ExPaSy tools.

PREDICTION OF SUB CELLULAR LOCALIZATION OF PROTEINS

The TBpred publically available online tool was used in this study. The models were trained and tested on 852 mycobacterial proteins and evaluated using five-fold cross-validation technique. A support vector machine (SVM) model using amino acid composition and overall accuracy of 82.51% with average accuracy of 68.47% was achieved. In order to utilize evolutionary information, a SVM model using PSSM profiles obtained from PSI-BLAST and overall accuracy achieved was 86.62% with average accuracy of 73.71%. In addition, HMM (Hidden Markov Model), MEME/MAST and hybrid model that combined two or more models were also developed. Overall accuracy of 86.8% with average accuracy of 89.00% using combination of PSSM based SVM model and MEME/MAST. The performance of this method was compared with that of existing methods developed for predicting subcellular locations of Gram-positive bacterial proteins (Rashid et al 2007).

RESULTS AND DISCUSSION

In this study we have selected thirty-nine putative protein of M. tuberculosis H37Rv and their nature was analyzed theoretically. The molecular weight and isoelectric point of all these proteins was deduced (Table 1). The pI value of culture filtrate protein ranged from 4.39 to 6.98, cell surface protein ranged from 4.18 to 8.87, the protein from lipid and fatty acid metabolism ranged from 4.96 to 8.53, amino acid & purine biosynthesis genes ranged from 4.99 to 6.24, metal uptake proteins ranged from 5.22 to 8.34, anaerobic respiration & oxidative stress protein ranged from 4.33 to 9.93, sigma factors proteins ranged from 4.26 to 6.62, response regulator proteins ranged from 4.84 to 6.22 and transcriptional factors protein ranged from 8.56 to 10.35. It indicates protein stability in a particular isoelectric point (pI). An online TBpred server was used to predict protein localization within bacteria or targeting the host. We investigate whole putative proteome of Mycobacterium and their specific subcellular location (Table 1).
An extent of utilization of human cellular localization mechanisms by bacterial proteins and that appropriate subcellular localization predictors can be used to predict bacterial protein localization within the host cell. This is a pathogenic strain of human. Therefore, we have selected secretory proteins, which is responsible for causing human disease. In this study, the subcellular localization of proteins within the M. tuberculosis was out of 39 proteins, the 24 proteins was cytoplasmic, 11 integral membrane, 2 secretory and 2 protein attached to membrane by Lipid anchor.

Previously, all the study has been done on the basis of in vitro assay for identification of subcellular localization of proteins. Since, there was no bioinformatics tool available to predict the location of protein for targeting the vaccine or targeting the drugs.

Very few reports were available on localization of proteins in vitro experiments. Fifty-two proteins of TM1 subgroup-only 7 had been previously reported to be secreted proteins. The predictions were confirmed in 9 out of 10 TM1 identified secreted proteins of M. tuberculosis with high efficiency and 90% accuracy (Gomez et al 2000). An additional, study have been done and reported seven novel antigens of M. tuberculosis, which previously identified were based on reactivity to sera isolated from patients infected with tuberculosis were characterized. The seven proteins studied and identified location. Two are cytosolic, while two others, had high proline contents, and were tightly associated with cell wall; one was secretory protein (Amara et al 1998).

In conclusion, we include the specified prediction of subcellular localization results in the most putative proteins of isolate M. tuberculosis H37Rv. This initiative might be useful in annotating newly sequenced or hypothetical mycobacterial proteins. Thus the search for a potential vaccine/ drug target for an important bacterial pathogen by in vitro researchers will greatly be appended by this prediction.

References
In silico analysis of subcellular localization of putative proteins of Mycobacterium tuberculosis H37Rv strain

Author Information

Pallavi Somvanshi
Biotech Park

Vijai Singh
Aquatic Microbes Section, National Bureau of Fish Genetic Resources

Prahlad Kishore Seth
Biotech Park