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

P Somvanshi, V Singh, P Seth

Citation

P Somvanshi, V Singh, P Seth. *In silico analysis of subcellular localization of putative proteins of Mycobacterium tuberculosis H37Rv strain*. The Internet Journal of Health. 2007 Volume 7 Number 1.

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 involve 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, secretary and membrane- attached proteins (Bendtsen et al 2004). No bioinformatics tool was present for predicting sub cellular 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, secretary 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 sub cellular 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 isoeletric 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 (Table1).

Figure 1

Table 1: Physico-chemical properties and subcellular localization of putative proteins of H37Rv.

S.No	Protein designation	Accession no.	Length of sequence (a.a)	Molecular weight (Da)	pI	Subcellular localization
1	Culture filtrate	NP_216547.1	144	16229.2	4.90	Cytoplasmic Protein
2	Culture filtrate	ABD98028.1	95	9905.1	4.39	Secreted protein
3	Culture filtrate	NP_218280.1	159	15148.9	6.98	Protein attached to membrane by Lipid anchor.
4	Culture filtrate	CAB06231.1	1718	165963.3	5.78	Cytoplasmic Protein
5	Cell surface	NP_218327.1	284	27704.2	4.18	Integral Membrane protein
6	Cell surface	NP_217456.1	2111	224407.6	4.98	Cytoplasmic Protein
7	Cell surface	CAA98985.2	583	63051.7	6.07	Cytoplasmic Protein
8	Cell surface	NP_217458.1	920	62648.2	5.16	Integral Membrane protein
9	Cell surface	NP_218321.1	338	95133.9	8.87	Secreted protein
10	Cell surface	NP_215156.1	301	35690.6	6.38	Cytoplasmic Protein
11	Cell surface	YP_177730.1	287	34674.3	5.52	Cytoplasmic Protein
12	Cell surface	NP_215414.1	326	33031.6	6.27	Integral Membrane protein
13	Cell surface	NP_214989.1	199	33578.3	7.28	Cytoplasmic Protein
14	Cell surface	CAE30332.2	96	26729.8	8.23	Integral Membrane protein
15	Cell surface	CAB06100.1	580	10730.5	8.79	Cytoplasmic Protein
16	Lipid and fatty acid metabolism	YP_001281756	428	47092.2	4.96	Cytoplasmic Protein
17	Lipid and fatty acid metabolism	NP_218004.1	277	29433.2	7.49	Integral Membrane protein
18	Lipid and fatty acid metabolism	P65211.	189	20030.3	8.53	Integral Membrane protein
19	Amino acid & purine biosynthesis genes	CAA16072.1	198	21782.4	4.99	Integral Membrane protein
20	Amino acid & purine biosynthesis genes	CAA94261.1	370	37744.6	6.24	Integral Membrane protein
21	Amino acid & purine biosynthesis genes	CAB00926.1	295	30175.5	4.58	Integral Membrane protein
22	Amino acid & purine biosynthesis genes	CAB02370.1	297	32934.0	5.03	Cytoplasmic Protein
23	Metal uptake	NP_216327.1	234	24809.6	8.34	Integral Membrane protein
24	Metal uptake	NP_216899.1	1414	151652.4	5.34	Cytoplasmic Protein
25	Metal uptake	NP_217227.1	230	25236.0	5.22	Cytoplasmic Protein
26	Anaerobic respiration & oxidative stress	NP_216252.1	652	72833.9	9.23	Integral Membrane protein
27	Anaerobic respiration & oxidative stress	AAY54558.1	740	80578.4	4.93	Cytoplasmic Protein
28	Anaerobic respiration & oxidative stress	YP_001283786.1	195	21568.7	4.33	Cytoplasmic Protein
29	Anaerobic respiration & oxidative stress	AAD49325.2	137	23036.5	6.24	Cytoplasmic Protein
30	Anaerobic respiration & oxidative stress	NP_214946.1	240	23847.7 Motif RGD 165	6.24	Protein attached to membrane by lipid anchor
31	Sigma factors	CAB09733.1	370	40987.5	6.62	Cytoplasmic Protein
32	Sigma factors	NP_218204.1	122	12652.2	5.59	Integral Membrane protein
33	Sigma factors	CAB10918.1	257	28880.5	4.95	Cytoplasmic Protein
34	Sigma factors	AAD49325.2	137	14306.9	4.26	Cytoplasmic Protein
35	Response regulators	NP_215271.1	247	27516.8	6.20	Cytoplasmic Protein
36	Response regulators	NP_215418.1	236	25257.0	4.84	Cytoplasmic Protein
37	Response regulators	pdb|3B4T	262	27603.6	5.46	Cytoplasmic Protein
38	Transcriptional factor I	NP_214867.1	126	14127.9	10.35	Cytoplasmic Protein
39	Transcriptional factor II	NP_217933.1	102	11613.5	8.56	Cytoplasmic Protein

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 secretary 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 secretary 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

1. Amara RR, Shanti S, Satchidanandam V 1998. Characterization of novel immunodominant antigens of Mycobacterium tuberculosis. Microbiology. 144 (5):1197-203.
2. Bendtsen JD, Jensen LJ, Bloom N, Von Heijne G, Brunak S: Featurebased prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel 2004, 17:349-356.
3. Bhasin M, Garg A, Raghava GP: PSLpred: prediction of subcellular localization of bacterial proteins. Bioinformatics 2005, 21:2522-252418-26.
4. Bhasin M, Raghava GP: ESLpred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST. Nucleic Acids Research 2004, 32:W414-W419.

5. Bloom, B. R., and J. D. Mckinney. 1999. The death and resurrection of tuberculosis. Nat. Med. 5:872-874.
6. Bloom, R. B., and C. J. L. Murray. 1992. Tuberculosis: commentary on a reemergent killer. Science 257:1055-1064.
7. Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FS: PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics 2005, 21:617-623.
8. Gomez M, Johnson S, Gennaro ML (2000). Identification of secreted proteins of Mycobacterium tuberculosis by a bioinformatics approach. Infect Immun. 68(4):2323-7.
9. Guo J, Lin Y: TSSub: eukaryotic protein subcellular localization by extracting features from profiles. Bioinformatics 2006, 22:1784-5.
10. Lodish H, Baltimore D, Berk A, Zipursky SL, Matsudaira P, Darnell J: Molecular Cell Biology 3rd edition. Scientific American Books, New York; 1995:739-777
11. Mamoon Rashid, Sudipto Saha and Gajendra PS Raghava (2007). Support Vector Machine-based method for predicting subcellular localization of mycobacterial proteins using evolutionary information and motifs. BMC Bioinformatics. 8: 337.
12. Shen HB, Yang J, Chou KC: 2007. Euk-PLoc: an ensemble classifier for large-scale eukaryotic protein subcellular location prediction. Amino Acids, 33:57-67.
13. Smith, I. (2003). Mycobacterium tuberculosis Pathogenesis and Molecular Determinants of Virulence. Clinical Microbiology Reviews, 16(3): 463-496
14. Xie D, Li A, Wang M, Fan Z, Feng H: LOCSVMPSI: a web server for subcellular localization of eukaryotic proteins using SVM and profile of PSI-BLAST. Nucleic Acids Research 2005, 33:W105-W110.

Author Information

Pallavi Somvanshi
Biotech Park

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

Prahlad Kishore Seth
Biotech Park