Journal of Advanced Bioinformatics Applications and Research ISSN 0976-2604
Vol 3, Issue 1, 2012, pp 246-249 http://www.bipublication.com

COMPUTATIONAL ANALYSIS OF SUBCELLULAR LOCALIZATION OF HYPOTHETICAL PROTEINS OF MYCOBACTERIUM TUBERCULOSIS H37RA STRAIN

Lakshmi Pillai, Usha Chauhan

Department Of Mathematics, MANIT, BHOPAL, M.P, INDIA
Corresponding author-lakshmi pillai, Email: lakshmilster@gmail.com

ABSTRACT

The Tuberculosis is the classical human mycobacterial disease, caused by *Mycobacterium Tuberculosis. 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. Thus there arises the need to understand various parameters of these proteins for prediction of their functionality. The computational approaches for prediction of their classes are fast and economical therefore can be used to complement the existing wet lab techniques. Realizing their importance, in this paper an attempt has been made for the *insilico* prediction of protein subcellular localization. As in the case of *Mycobacterium,* proteins are often involved in extensive interactions at various subcellular localizations in cell. Total one thousand four hundred and thirty-two hypothetical 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. And they have widespread applications in function of proteins in the host cell and in designing the tuberculosis drugs.

Keywords: Sub cellular localization, pathogenicity, chemotherapeutic drugs, virulence factors, hemolytic molecules, secretary proteins.

INTRODUCTION

Mycobacterium tuberculosis continues to be the major infectious cause of human death in developing countries and has reemerged in industrialized countries. 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[1], without treatment up to 60% people infected will be dying. Its major rationales were poverty, lack of healthy living conditions and adequate medical care. Tuberculosis continues to affect about 30% world's population, mainly in developing countries, despite existence of chemotherapeutic drugs and widespread use of the *Mycobacterium Bovis* BCG vaccine. Effective chemotherapeutic treatment is expensive, takes long time and not available to people in various parts of world. 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 [2].

Various methods had been developed for predicting sub cellular location of eukaryotic, prokaryotic (Gram-negative and Gram-positive bacteria) but only one method has been developed for mycobacterium protein. In this analysis, an attempt is made to predict 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 [3]. Earlier, cellular localization of *M. tuberculosis* is based on *in vitro* assay like ELISA, western blotting and *in situ*. 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.

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, LOCSVMPSI [4], TSSub [5], ESLpred [6], Euk-Ploc [7]. As various tools were available for prediction of subcellular localization of prokaryotic proteins viz. PSORTb [8], PSLpred [9]. A model has been developed for predicting four subcellular locations of mycobacterium proteins, namely cytoplasmic, integral membrane, secretary and membrane- attached proteins. In the present study, 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 is used.

The method used here was an indirect method where attempt have been made to predict subcellular localization of proteins rather than function. The subcellular localization methods are based on observation that protein belongs to same compartment of protein has similar amino acid composition [10] and has similar functions.

The aim behind this study was to predict the sub cellular localization of putative proteins of *M. tuberculosis* H37RA strain as they might be useful for targeting antimycobacterial drugs.

MATERIALS AND METHODS

Collection of sequences

The complete protein sequences of cell surface, lipid & fat metabolism, amino acid & purine biosynthesis genes, anaerobic respiration & oxidative stress, metal uptake of *Mycobacterium tuberculosis* H37RA were extracted from biological database National Centre for Biotechnology Information (NCBI) cited at http://www.ncbi.nlm.nih.gov

Prediction of sub cellular localization of proteins

The PSLPred publically available online tool was used in this study. PSLPred is a SVM based method, predicts 5 major subcellular localization (cyotplasm, inner-membrane, outer-membrane, extracellular, periplasm) of Gram-negative bacteria. This method includes various SVM modules based on different features of the proteins such as - Amino acid composition, Dipeptide Composition, Composition of 33 physico-chemical properties, and evolutionary information of PSI-Blast. The overall prediction accuracy of these SVM module is 86%, 86%, 83% and 68% respectively. In addition, hybrid approach based SVM module has also been developed based on all above mentioned features of a protein. In this paper subcellular localization s predicted based on amino acid composition with an accuracy of 86%.

RESULTS AND DISCUSSION

In this study we have selected one thousand four hundred and thirty two putative proteins of *M. tuberculosis* H37RA and subcellular localization was analyzed.

An online PSLPred 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).

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 [11], which is responsible for causing human disease. In this study, the subcellular localization of proteins within the *M. tuberculosis* was out of 1432 proteins (Table 2), the 1417 proteins was cytoplasmic, 13 integral membrane, 1 secretory and 1 protein attached to membrane by Lipid.

CONCLUSION

Previously, all the study has been done on the basis of in vitro assay for identification of subcellular localization and function of proteins. Since, there was no computational 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. In conclusion, we include the specified prediction of subcellular localization prediction results in the most putative proteins of *M. tuberculosis* H37RA. 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.

ACKNOWLEDGMENT

The authors are highly thankful to the Department of Biotechnology, Delhi, India for providing support in the form of Bioinfomatics infrastructure facility to carry out this work.

REFERENCES

1. Bloom B. R. and J. D. Mckinney 1999, The death and resurrection of tuberculosis *Nat. Med.* 5:872-874 [PMID: 10426305].

2. Bloom, B.R and C. J. Murray 1992, Tuberculosis: commentary on a reemergent killer *Science* 257:1055-1064 [PMID: 1509256]

3. Lodish H et al. 1995, *Molecular Cell Biology* 3rd edition, Scientific American Books, New York: 739-777.

4. Xie D et al. 2005, LOCSVMPSI: a web server for subcellular localization of eukaryotic proteins using SVM and profile of PSI-BLAST *Nucleic Acids Research*, **33:**W105-W110 [PMID: 15980436].

5. Guo J and Lin Y 2006, TSSub: eukaryotic protein subcellular localization by extracting features from profiles *Bioinformatics* , **22:**1784-5 [PMID: 16787975].

6. Bhasin M and Raghava GP, ESLpred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST **32:**W414-W419 [PMID: 15215421].

7. Shen HB et al. 2007, Euk-PLoc: an ensemble classifier for large-scale eukaryotic protein subcellular location prediction *Amino Acids*, **33:**57-67 [PMID: 17235453].

8. Gardy JL et al. *Bioinformatics* 2005, PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. 21:617-623 [PMID: 15501914].

9. Bhasin M et al. 2005, PSLpred: prediction of subcellular localization of bacterial proteins *Bioinformatics*, 21:2522-252418-26[PMID: 15699023].

10. Rost B et al. 2003 Dec; Automatic prediction of protein function Cell Mol Life Sci. 60(12):2637-50. Review [PMID: 14685688].

11. Gomez M et al. 2000, Identification of secreted proteins of Mycobacterium tuberculosis by a bioinformatics approach. 68(4):2323-7 [PMID: 10722636].

Table 1: Subcellular localization of putative proteins of *Mycobacterium tuberculosis* H37RA.

Acc id	amino acid based subcellular prediction
ZP_02553214	CYTOPLASMIC PROTEIN
ZP_02553212.1	INTEGRAL MEMBRANE PROTEIN
ZP_02553209.1	INTEGRAL MEMBRANE PROTEIN
ZP_02553208.1	CYTOPLASMIC PROTEIN
ZP_02553207.1	INTEGRAL MEMBRANE PROTEIN
ZP_02553206.1	CYTOPLASMIC PROTEIN
ZP_02553205.1	INTEGRAL MEMBRANE PROTEIN
ZP_02553204.1	SECRETED PROTEIN
ZP_02550120.1	CYTOPLASMIC PROTEIN
ZP_02549605.1	CYTOPLASMIC PROTEIN
ZP_02549604.1	CYTOPLASMIC PROTEIN
ZP_02549603.1	PROTEIN ATTACHED TO MEMBRANE BY LIPID ANCHOR
ZP_02549602.1	INTEGRAL MEMBRANE PROTEIN
ABR14062.1	CYTOPLASMIC PROTEIN
ABR14061.1	CYTOPLASMIC PROTEIN
AAB07556.1	CYTOPLASMIC PROTEIN
CAB05953.1	INTEGRAL MEMBRANE PROTEIN
ZP_02553232.1	INTEGRAL MEMBRANE PROTEIN
ZP_02553231.1	CYTOPLASMIC PROTEIN
ZP_02553230.1	INTEGRAL MEMBRANE PROTEIN
ZP_02553229.1	INTEGRAL MEMBRANE PROTEIN
ZP_02553228.1	CYTOPLASMIC PROTEIN
ZP_02553227.1	CYTOPLASMIC PROTEIN
ZP_02553226.1	CYTOPLASMIC PROTEIN
ZP_02553225.1	CYTOPLASMIC PROTEIN
ZP_02553224.1	CYTOPLASMIC PROTEIN
ZP_02553223.1	INTEGRAL MEMBRANE PROTEIN
ZP_02553222.1	CYTOPLASMIC PROTEIN
ZP_02553221.1	CYTOPLASMIC PROTEIN
ZP_02553220.1	CYTOPLASMIC PROTEIN

Table 2: Subcellular localization Predicted

Protein Locations	**Secretory Proteins**	**Cytoplasmic proteins**	**integral membrane protein**	**protein attached to membrane by Lipid**
Number of protein in particular location	1	1417	13	1