 Proceedings of ICCES'08, pp.438-442

in silico Method for the Identification of *Mycobacterial* sp. Potential Drug Targets

Ashutosh Singh[1], Shruti Mishra[2], Dhwani Raghav[1]
Asheesh shanker[1] and Vinay Sharma[1]

Summary

Drug resistance has increased the pace of war against the ever-growing challenge of mycobacterial infections particularly with the growing menace of tuberculosis (TB).Previous studies reported several essential and virulent genes of mycobacterium like virS gene and mymA operon[1] through experimental approaches. However, Post genomic approach applied for the identification of targets for tuberculosis which includes the comparison of *Mycobacterium tuberculosis* CDC1551 proteome against database of essential genes and proteome of *Homo sapiens*. A total of approx 4000 proteins were studied and compared and 19 proteins were found to possess potentiality to call as Targets.

Introduction

For centuries, tuberculosis has accounted more human misery, suffering, loss of earnings, and failure of economic and social development than any other disease. According to World Health Organization TB is the world's longest running catastrophe, killing more than 200 people every hour and more than 5000 day [4]. However, the complexity of pathogen *(M.tuberculosis)* [3] and host (human) genome [5] is solved in fact some of the strains of the *Mycobacterium tuberculosis* has been re-annotated [2] still no new drug has been invented in past 50 years. For over 40 years, isoniazid (isonicotinic acid hydrazide (INH)) has been utilized as a frontline agent in drug mixtures to treat *Mycobacterium tuberculosis*. The tubercle bacillus owes its virulence to its ability to survive within the macrophage; the mechanism of virulence is poorly understood and is supposed to be multifactorial.The presence of significant sequence diversity among mycobacterial species would provide a basis for understanding differences in their virulence and a method to identify essential genes which are responsible for pathogenesis and hence the proteome of *Mycobacterium tuberculosis* intensively studied for identification of proteins or enzymes that could serve as targets for tuberculosis.

Essentiality was also determined by sequence similarity approaches.

Method

Targets were selected through genomics - related methodologies. The genes that are present in the genome of the *M.tuberculosis* but not in the genome of the

[1]Faculty of Life Sciences, Department of Bio techgnology, Banasthali University, Banasthali, Rajasthan, India. Email: ashutosh.bio@gmail.com
[2]Research Associate, Biobase, Banglore, India.

closely related free living bacterium as well as in human are therefore likely to be important for pathogenicity and may be considered as candidate drug targets. Moreover, the genes necessary for the proper functioning of the pathogenic bacteria are essential genes, and can be treated as drug targets if they have no homolog in *Homo sapiens*, these candidate drug targets were selected after determining the essentiality in the pathogen.

Complete genome of *M.tuberculosis CDC1551* of accession no. NC_002755 exploited and proteome sequence downloaded from the NCBI ftp server. The essential genes database (DEG) [6] then used for comparison against the *M.tuberculosis* protein sequences. The Hits found were saved for the next level of analysis. An E-value of less than 1 and positivity greater than or equal to 60% was used as a cut-off for defining a biologically relevant relationship. The reason for choosing higher E-value for the DEG blast, as higher E-Value generates more number of hits with different functions, if more number of hits found to share same function irrespective of the lower score for some, would increase there probability as essential genes.

BLASTP was also performed against *Homo sapiens* to map essential genes on human proteome. The non homologous essential genes were considered as targets for the pathogen. This has greatly reduced the number of probable essential genes identified. During this analysis, only 19 sequences, which showed no hits in a comparison against human proteome, selected and are the essential genes having capabilities to be called potential drug targets. There were almost 10 unannotated proteins, their function yet to be determined. However, the remaining almost 10 proteins are annotated and can be positively defined as potential drug targets for *M.tuberculosis*.

Results

The BLASTP was performed on 4189 protein sequences of *Mycobacterium tuberculosis* against DEG. An E-value of less than 1 and positivity greater than 60 % was used as a selection criterion to identify the sequences for further analysis. The numbers of sequences identified were 352. These 352 sequences were subjected to BLASTP against *Homo sapiens* to identify non-homologous sequences which may serve as potential drug targets. 19 essential genes have been identified in M. tuberculosis by a computational genomic approach. **Figure 1** illustrates the distribution of hits obtained in BLASTP comparison results:

These proteins were further classified according to the function they performed which is listed in the database. Rest 10 was found to be hypothetical proteins whose putative functions were identified by comparing them to the homologous sequences retrieved from the BLASTP search. It should be emphasized that our

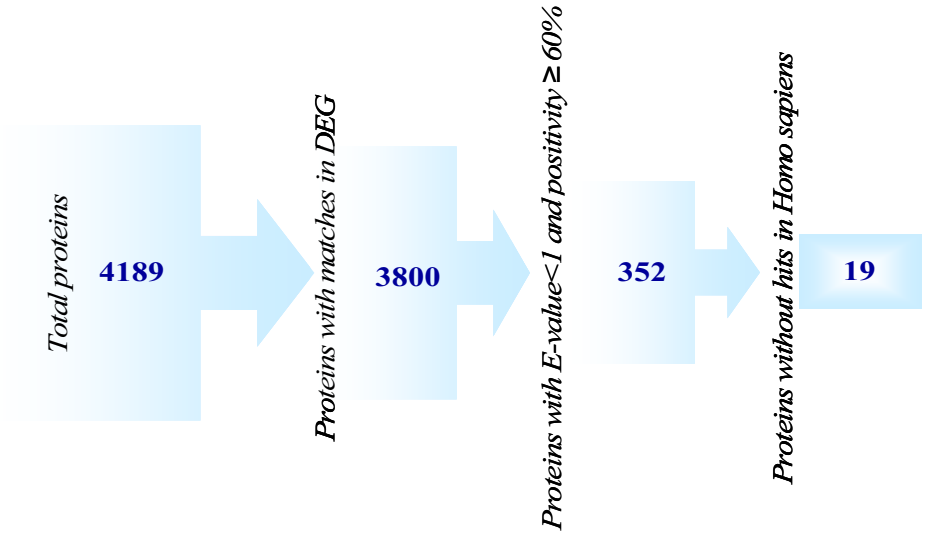

Figure 1: Summary of the distribution of hits obtained in BLASTP

analysis excluded proteins that were not conserved in other related genera and also homologous to human genome. **Table 1** lists functions of all the 19 essential genes.

The complete genome sequence of the tuberculosis bacillus, *Mycobacterium tuberculosis*, was presented in *Nature* in June 1998. Although directly observed short course chemotherapy exists to treat the disease, the emergence of drug-resistant strains has severely threatened the efficacy of the treatment. The sequencing of the *M. tuberculosis* genome holds promise for the development of new vaccines and the design of new drugs. This is all the more possible when the information from the genome sequence is combined with proteomics and structural and functional genomics.

This study reveals that though the drug targets identified, the definition of the half of the proteins are yet to be determined. Additionally, only set of proteins were considered as the proteins with no ortholog in DEG were rejected however It is probable that many of these proteins play essential roles and therefore represent attractive drug targets but due to absence of their ortholog from the DEG they were not considered. To elucidate the structures of the potential drug targets or unknown proteins belonging to large gene families and apply rational drug design approach to dock the potential drug targets by screening with the libraries of natural or chemical compounds demands more study. Screening of chemical and natural product libraries will identify potential inhibitors that could correspond to lead compounds for the development of new drugs to treat tuberculosis. Besides that future work will be on novel targets and pathways for the development of antibacterial agents. The protein-protein interactions will in turn serve as the basis of target-oriented

Table 1: Non-homologous protein sequences in *Homo sapiens*

Acc. No.	Gene	Function	Putative function
NP_334468.1	ssb	single-strand binding protein	
NP_334782.1	MT0375	hypothetical protein MT0375	Probable conserved integral membrane protein
			Zn-dependent proteases
			putative peptidase
NP_335091.1	rplJ	ribosomal protein L10	
NP_335207.1	MT0780	hypothetical protein MT0780	putative transposase
			Transposase and inactivated derivatives
NP_335268.1	MT0840	transcriptional regulator	
NP_335958.1	MT1508	conserved hypothetical protein	
NP_336055.1	MT1602	hypothetical protein MT1602	ribosomal protein L1
			LSU ribosomal protein L1P
NP_336554.1	MT2089	hypothetical protein MT2089	Erythromycin esterase homolog
NP_337108.1	aroQ	3-dehydroquinate dehydratase	
NP_337164.1	yajC	protein-export membrane protein	
NP_337315.1	MT2808	hypothetical protein	Conserved Hypothetical Arginine Rich Protein
NP_337595.1	MT3080.1	hypothetical protein	Possible Exported Protein
NP_338067.1	MT3541	hypothetical protein MT3541	Uncharacterized iron-regulated membrane protein
NP_338108.1	MT3573.8	hypothetical protein MT3573.8	phage prohead protease, putative
			Phage head maturation protease
			possible phage protease, R. capsulatus GTA orfg4 homologue
NP_338374.1	MT3819	hypothetical protein MT3819	Uncharacterized protein conserved in bacteria
NP_338466.1	MT3914	PAP2 superfamily protein	
NP_338475.1	MT3923	acyltransferase family	
NP_338581.1	sigM	RNA polymerase sigma-70, ECF subfamily	
NP_338250.1	panD	aspartate 1-decarboxylase	

drug screening, as all know how much epidemic and fatal tuberculosis is, and as the sequence mystery already solved we can now do more on the drug discovery that would be more fast, accurate and cost effective.

References

1. Ramandeep Singh,Amit Singh,Anil K Tyagi (2005). Deciphering the genes involved in pathogenesis of *Mycobacterium tuberculosis.Tuberculosis*. 85, 325–335.
2. Camus, J. C., M. J. Pryor, C. Medique and S. T. Cole (2002). Reannotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology*. 148: 2967–2973.
3. Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas and C. E. Barry (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*. 393: 537–544.
4. C. Dye, S. Scheele, P. Dolin, V. Pathania and M.C. Raviglione. (1999). *JAMA* 282 , 677–686.
5. J. Craig Venter *et al.* (2001). The Sequence of the Human Genome.*Science*. 291: 5507, 1304–1351.
6. Ren Zhang, Hong-Yu Ou and Chun-Ting Zhang (2003). DEG: a database of essential genes.*Nucleic Acids Research*. 32 D271–D272, DOI: 10.1093/nar/g-kh024.