IJPSR (2011), Vol. 2, Issue 7 **(Research Article)**

INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES AND RESEARCH

Received on 19 April, 2011; received in revised form 09 May, 2011; accepted 18 June, 2011

SUBTRACTIVE GENOMICS APPROACH FOR *IN SILICO* IDENTIFICATION *OF* NOVEL DRUG TARGETS AND EPITOPES FOR VACCINE DESIGN IN *TREPONEMA PALLIDUM SUBSP. PALLIDUM STR. NICHOLS*

Vijayakumari Malipatil [1], Shivkumar Madagi [1] and Biplab Bhattacharjee*[2]

DBT BIF Center, Karnataka State Women University [1], Bijapur, Karnataka, INDIA
Institute of Computational Biology (IOCB) [2], Bangalore, Karnataka, India

Keywords:

Subtractive genomics,
T. pallidum SS14,
Novel drug targets,
Syphilis,
Essential genes,
Putative drug targets,
Membrane proteins

Correspondence to Author:

Biplab Bhattacharjee
Senior Scientist, Institute Of Computational Biology, Domlur Layout, Bengaluru-560071, Karnataka, India

ABSTRACT

In silico differential genomics helps to identify genes that encode for unique metabolism with relation to human. The genomic database provides a vast amount of useful information for the drug target identification. The subtractive dataset obtained comparatively between the human and the pathogen genome, differentially provides information about the genes that are likely to be essential to the pathogen but is not part of the host (human).This approach has given fruitful results in recent times to identify essential genes in *Pseudomonas aeruginosa.* The same strategy is used to analyse the whole genome sequence of the *Treponema pallidum subsp. pallidum str. Nichols*. Three putative membrane-bound drug targets have been derived step-wise, out of the 301 essential genes that have been predicted. The putative drug targets include the drug targets taking part in unique metabolic pathways that are situated in the membrane and are specific to the pathogen. Structure prediction of the membrane bound drug targets is done along with B-cell epitope mapping that highlights the immunogenic part of a protein. Syphilis is characterised by many asymptomatic and latent clinical stages. In spite of effective prophylaxis by use of penicillin, there has been increase in the resistance in the pathogen and an alternative is required due to penicillin allergic pregnant women. *In silico* study for identification of potential drug targets has been possible due availability of whole proteomic data of *Treponema pallidum subsp. Pallidum str. Nichols*.

INTRODUCTION: *Treponema pallidum subsp. pallidum str. Nichols* is one of the strains that cause venereal disease syphilis. This was isolated originally from neurosyphilitic patient. Others strains like *pallidum* str p*ertenue*, *carateum* and *endemicum* cause the skin infections yaws, pinta and bejel, respectively. *Treponema pallidum subsp. pallidum str. Nichols* is slender helical shaped gram negative bacteria that have an unusual cell envelope when compared to other gram negative bacteria.

Pathogenicity of the bacteria is mainly due to the presence of the capsule. It consists of the outer and inner membrane that helps in studying the membrane associated proteins [1]. This organism is the causative agent of venereal syphilis at the molecular level. The sexual transmitted disease was first discovered in Europe at the end of the fifteenth century, however, the causative agent was not identified until 1905 [2]. Syphilis was reported to be the third most commonly reported transmittable disease in USA.

Syphilis is characterized by many clinical stages and long periods symptomless and latent infection. Although effective chemotherapies have been available as a result of use of penicillin, syphilis remains a major global health problem.

Treponema pallidum subsp. pallidum str. Nichols shows striking similarity with *E. coli* ribosomal proteins that confer microlide resistance [3]. Resistance can be overcome by alternative drug targets as well as alternative drugs. Complete genomic sequence and proteomic data is available due large scale sequencing projects in the public domain [4].

MATERIALS AND METHODS: The subtractive genomics methodology retains the protein dataset that is indispensable and the proteins that are part of the unique metabolic pathway. The proteins essential for the basic function of *Treponema pallidum subsp. pallidum str. Nichols* are analysed further for structure prediction and epitope mapping. The flow chart in **figure 1** shows algorithm for the present approach [5, 6, 7, 8].

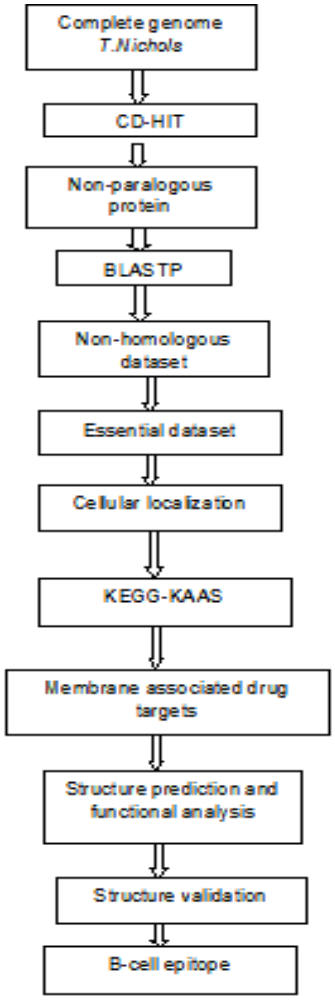

FIGURE 1: FLOWCHART FOR IDENTIFYING ESSENTIAL PROTEINS

Retrieval of proteome of Host and Pathogen: The complete proteome and the NR datset of the *Treponema pallidum subsp. pallidum str. Nichols* and human were retrieved from NCBI [9]. Essential protein sequences of the pathogen are manually extracted from the database of essential genes (DEG) [10].

Identification of Essential Proteins in *T. Pallidum* str Nichols: Paralogs are removed from *Treponema pallidum* subsp. pallidum str. *Nichols proteome by using CD-HIT set at 60% threshold* [11]. The non-paralogous proteins obtained were subjected to sequence similarity with proteins of *Homo sapiens* by BLASTp with the expectation value (E-value) cutoff of 60% sequence identity. The protein sequences of *Treponema pallidum* subsp. *pallidum* str. Nichols showing less significant similarity with *Homo sapiens* proteome was retrieved manually.

The non-homologous proteins of *Treponema pallidum* subsp. *pallidum* str. *Nichols* are studied for their essentiality by BLASTp against database of essential proteins E-value cut off score of 10-10. The program used to score the essential genes is DEG. The minimum bitscore threshold to screen out non-essential proteins is set at greater than 100. The resulting proteins are the non-homologous essential proteins of *Treponema pallidum subsp. pallidum* str. Nichols.

Protein function prediction of essential genes: The function of the uncharacterised proteins is predicted using SVMProt web server (http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi.) [12]. The primary sequence of Proteins is used to characterize proteins according to their function by using SVM (Support Vector machine) Prot program.

Metabolic Pathway Analysis: Essential proteins of *Treponema pallidum subsp. pallidum str. Nichols* was subjected to metabolic pathway analysis by KEGG Automatic Annotation Server (KAAS) [13]. The comparative analysis of the metabolic pathways of the of the host and the pathogen was performed using Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway database [14] to sort out essential proteins in the pathogen that mediate specific metabolic pathways for the identification of unique potential drug targets.

Sub Cellular Localization Prediction: Finding the cellular position of the subtractive protein-set was accomplished by CELLO (subcellular localization predictor) program [15] to identify the membrane-bound proteins which could be probable drug targets.

Structure prediction of Membrane Drug Targets: The 3-D structure of the membrane bound proteins of the *Treponema pallidum subsp. pallidum str. Nichols* was predicted by Pyre2 server [16], since its sequence similarity with the know PDB structure was very less.

B-cell epitope mapping: The membrane proteins of the *Treponema pallidum subsp. pallidum str. Nichols* are the ideal vaccine candidate for the peptide vaccine preparation. The B-cell epitopes of the proteins are predicted using BCPreds (cutoff score >. 7) under default condition and its exposure to the external of the cell is identified by TMHMM. The antigenicity of the protein was determined by the VaxiJen server under default conditions (cutoff score >. 4) [17].

RESULTS AND DISCUSSION: Of the total 1036 genes of *Treponema pallidum subsp. pallidum str. Nichols* 301 are the essential genes belonging to different protein classes. 3 proteins are membrane-bound having role in unique metabolic pathways.

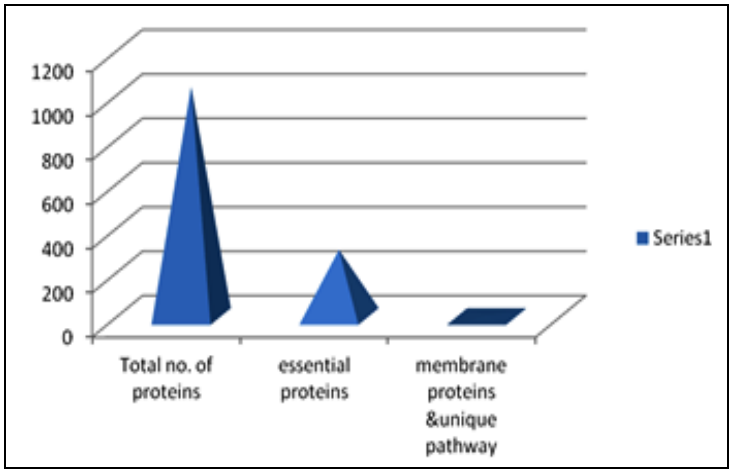

FIGURE 2: SUBTRACTIVE DATASET IN *TREPONEMA PALLIDUM* SUBSP. PALLIDUM STR. NICHOLS

The structural and functional proteins of *Treponemapallidum pallidum str. N*ichol showing more conserved regions with the human proteome are excluded from further analysis during execution of BLASTp. The resulting non-homologous proteins are analysed by BLASTp using another server DEG (database of essential genes) that predicts the essentiality of the gene. 301 essential genes are predicted from the proteome of the pathogen. The subtractive dataset is indicated in **figure 2**.

The sub cellular localization prediction of essential protein of *T. Pallidum str Nichols* are predicted to extract the proteins that are exclusively membrane-associated and also cross checking with the SVM prediction for the localization of the protein. Common results were included. Following this procedure a total of 3 proteins that had a high probability of being located in the membrane are considered.

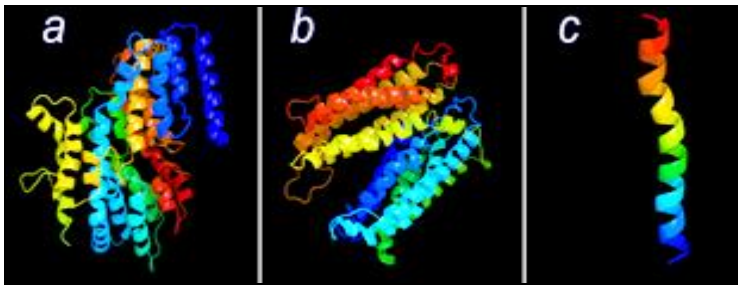

FIGURE 3: 3-DIMENSIONAL STRUCTURE OF THE MEMBRANE PROTEINS: a) DICARBOXYLATE TRANSPORTER (dctM), b) VIRULENCE FACTOR (mviN), c) CELL DIVISION PROTEIN (ftsW).

Analysis of the metabolic pathway of the membrane bound proteins in *Treponema pallidum subsp. pallidum str.* Is done by using KEGG Automatic Annotation Server (KAAS). A comparative analysis of the metabolic pathways predicted in pathogen to the human pathways is performed using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. Unique pathway noted during this analysis is Cell cycle – Caulobacter, Peptidoglycan biosynthesis, Two-component system.

Functional classification of the 3 putative uncharacterized membrane-bound essential proteins were performed by using the SVMProt web server. The proteins functions based on p value, which is the classification accuracy indicator. All Putative drug targets are predicted to have transmembranal function.

The 3-D structure of the proteins predicted through fold based method was validated by Ramchandran plot [18]. The percentage of amino acids of the predicted structure falling in of allowed regions of Ramchandran plot, the structural and functional analysis are summarized in **table 1**.

TABLE 1: FUNCTIONAL AND STRUCTURAL ANALYSIS OF THE PUTATIVE DRUG TARGETS IN RAMCHANDRAN PLOT

Name of protein	Gen identifier	Metabolic pathway	Protein function	% of amino acids in allowed region of Rammchandran plot
Dicarboxylate transporter (dctM)	15639942	Cell cycle - Caulobacter	Transmembrane	86.8%
Virulence factor (mviN)	15639507	Peptidoglycan biosynthesis	Transmembrane	64.6%
Cell division protein (ftsW)	15639378	Two-component system	Transmembrane	100.0%

B-cell epitope mapping showed that all membrane proteins are antigenic, especially the cell division protein. A total of 8 B-cell epitopes are predicted out of them only 3 epitope are exposed to the surface.

The virulence factor protein did not show any surface exposed epitopes instead they are transmembranal. Resulting parameters during epitope mapping is summarized in **table 2.**

TABLE 2: B-CELL EPITOPE MAPPING IN *TREPANOMA PALLIDUM* STR. NICHOLS

Name of protein	Gen identifier	Exposed B-cell epitope	Antigenicity score	position
Dicarboxylate transporter(dctM)	15639942	PLAVHFGVHPVHASVVFLMN	0.4493	556
Virulence factor (mviN)	15639507	-----	0.4385	---
Cell division protein (ftsW)	15639378	RGIGNGVRKIASVPEVYSDF VPATGIPLPFFSSGGSSIVV	0.5941	253 338

CONCLUSION: Huge amount of proteomic and genomic data is available in the public domain due to large scale genomic projects. DEG is an efficient tool for identification of drug targets in the genomic data under study [19]. In the current study, several proteins are studied that can be effective drug targets and posses epitopes for drug design & and vaccine design respectively in *Treponema pallidum subsp. pallidum str. Nichols*. Such drug targets will specifically "kill" the pathogens since drugs dock with proteins *in vivo* that are part of unique biochemical pathway. The essential genes of the *Treponema pallidum subsp. pallidum str. Nichols,* identified in the present study can be further studied for immunogenic portions of proteins. Virtual screening also helps in identifying the compound acting against these proteins.

ACKNOWLEDGMENTS: The authors are grateful to DBT-BIF Center, Karnataka State Women University, Bijapur for providing a platform for the research work.

Competing Interests: The authors declare that they have no competing interests.

REFERENCES:

1. Liu J, Howell JK, Bradley SD, Zheng Y, Zhou ZH, Norris SJ: Cellular architecture of *Treponema pallidum*: novel flagellum, periplasmic cone, and cell envelope as revealed by cryo electron tomography. J Mol Biol. 2010; 403(4):546-61.
2. Antal GM, Lukehart SA and Meheus AZ: The endemic treponematoses. Microbes Infect. 2002; 4 (1): 83–94.
3. L. V. Stamm and H. L. Bergen: A Point Mutation Associated with Bacterial Macrolide Resistance Is Present in Both 23S rRNA Genes of an Erythromycin-Resistant *Treponema pallidum* Clinical Isolate. Antimicrob Agents Chemother. 2000; 44(3): 806–807.
4. http://www.ncbi.nlm.nih.gov/
5. Allsop, A. E: New antibiotic discovery, novel screens, novel targets and impact of microbial genomics. Curr. Opin. Microbiol.1998; 1: 530-534.
6. Sakharkar KR, Sakharkar MK and Chow VT K: A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa. *In Silico* Biol. 2004; 4(0028):355.
7. Salama NR, Shepherd B, and Falkow S: Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*. J. Bacteriol. 2004; 186: 7926-7935.
8. Dutta A, Singh SK, Ghosh P, Mukherjee R, Mitter S and Bandyopadhyay D: *In silico* identification of potential therapeutic targets in the human pathogen *Helicobacter pylori*. *In Silico* Biol. 2006; 6(1-2) 43-7.
9. Fraser CM, Norris SJ, Weinstock GM, et al: Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. Science (journal). 1998; 281 (5375): 375–88.
10. Zhang R, Ou HY and Zhang CT: DEG: A database of essential genes. Nucleic Acids Res. 2004; 32: D271-D272.
11. Li W, Godzik A: Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics.2006; 22: 1658-1659.
12. Cai CZ, Han LY, Ji ZL, Chen X and Chen YZ: SVMProt: Web-Based Support Vector Machine Software for Functional Classification of a Protein from Its Primary Sequence. Nucleic Acids Res .2003; 31: 3692-3697.
13. Moriya Y, Itoh M, Okuda S, Yoshizawa A C, Kanehisa M: 2007 KAAS: an automatic genome annotation and

pathway reconstruction server; Nucleic Acids Res.2007; 35: 182.

14. Kanehisa M and Goto S: KEGG: Kyoto Encyclopaedia of Genes and Genomes. Nucleic Acids Res.2000; 28: 27.
15. Yu C S, Chen Y C, Lu C H, Hwang J K: Prediction of protein subcellular localization. Proteins: Structure, Function and Bioinformatics. 2006; 64: 643-651.
16. Kelley LA, *et al*.: Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc.2009; 4:363-371.
17. Debmalya Barh, Amarendra Narayan Misra, Anil Kumar, and Azevedo Vasco: A novel strategy of epitope design in *Neisseria gonorrhoeae*. Bioinformation. 2010; 5(2): 77–85.
18. Ramchandran G N, Ramakrishnan C, Sasisekharan V: Streochemistry of polypeptide chain configuration; J.Mol.Biol.1963; 7: 95-99.
19. Tomson FL, Conley PG, Norgard MV and HagmanKE: Assessment of cell-surface exposure and vaccinogenic potentials of *Treponema pallidum* candidate outer membrane proteins. Microbes Infect. 2007; 9 (11): 1267–75.

****