International Journal of Bioinformatics and Biological Science: v.1 n.1 p.79-85. March, 2013

Structure prediction of drug target identified by metabolic pathway analysis of *Streptococcus pyogenes*

Neetu Singh[1], Himanshi Kanojia[1], Satendra Singh[1], Deepak Kumar Verma[2], Budhayash Gautam[1] and Gulshan Wadhwa[*2]

[1]Department of Computational Biology & Bioinformatics, Jacob School of Biotechnology and Bioengineering, Sam Higginbottom Institute of Agriculture, Technology and Sciences-Deemed to be University, Allahabad-211007, India
[2]Department of Molecular and Cellular Engineering, JSBB, SHIATS, Allahabad-211007,India

[*]Corresponding Author: Gulshan Wadhwa gulshan@dbt.nic.in

Received: 24 December 2012; Accepted: 03 January 2013

ABSTRACT

Target identification is the first step in the drug and vaccine discovery process , *in silico* subtractive genomics is widely used in this process. We used this *in silico* approach for identification of essential proteins in *Streptococcus pyogenes* strain *M*1, a Gram-positive bacterium, which is an important human pathogen. *S. pyogenes* causes wide variety of disease, including pharyngitis (streptococcal sore throat), scarlet fever, impetigo, erysipelas, cellulitis, septicemia, toxic shock syndrome, necrotizing fasciitis (flesh-eating disease) and the sequelae, rheumatic fever and acute glomerulonephritis in humans. Further, using the KEGG Automatic Annotation Server, we identified six unique metabolic pathways that exist in this bacterium but not in human and some essential *S. pyogenes* proteins that are involved in these bacterial pathways. Analysis using CELLO, showed that essential *S. pyogenes* strain *M*1 proteins were membrane proteins and thus more amenable as drug targets. Therefore, based on identification, six drug target have been identified and their genes: SPy_1283, SPy_0890, SPy_0399, SPy_1849, SPy_1652, SPy _1233 whereas homology modeling for 6-phosphofructokinase were performed for inhibitor designing. This work has a lot of potential in research work for development of strategies for control and treatment of *S.pyogenes* , including targets that have been identified with this technique and various other aspects.

Keywords: Streptococcus pyogenes, *Metabolic pathway, Drug target*

BACKGROUND

Streptococcus pyogenes is a Gram-positive, nonmotile, nonperforming coccus that occurs in chains or in pairs of cells. Individual cells are round-to-ovoid cocci, 0.6-1.0 micrometer in diameter (Figure 1). The metabolism of *S. pyogenes* is

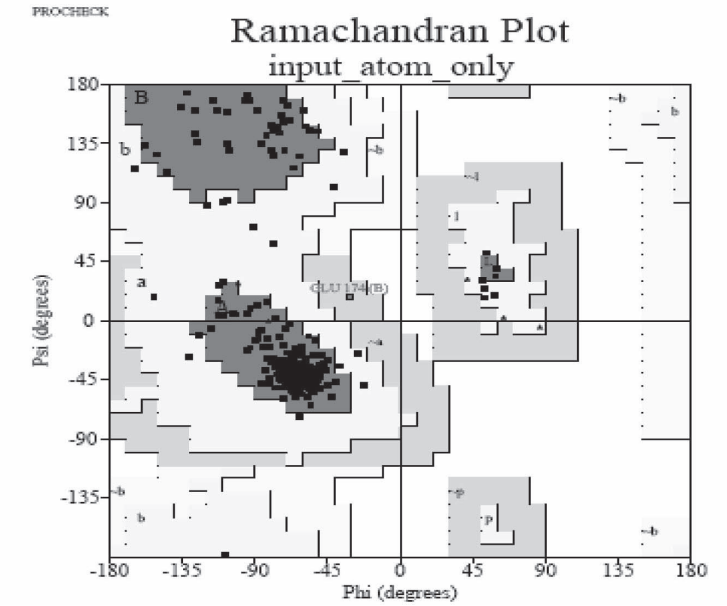

Figure 1: Ramachandran Plot

fermentative; the organism is a catalase-negative aerotolerant anaerobe (facultative anaerobe), and requires enriched medium containing blood in order to grow. Group A streptococci typically have a capsule composed of hyaluronic acid and exhibit beta (clear) hemolysis on blood agar(Rosenbach ,1884).

S. pyogenes is one of the most frequent pathogens of humans. However, there has been a recent increase in variety, severity and sequel of *S. pyogenes* infections, and a resurgence of severe invasive infections, prompting descriptions of "flesh eating bacteria" in the news media (Bisno and Stevens ,1996). Today, the pathogen is of major concern because of the occasional cases of rapidly progressive disease and because of the small risk of serious sequelae in untreated infections(Bisno 1991).

S. pyogenes lies among the Group A S., which is the leading cause of acute bacterial pharyngitis/tonsillitis, or "strep throat" occurring worldwide" (Caruana)(Falags *et al* ,2008) However, as the flesh-eating bacteria, it is not commonly seen, often not permitted to cumulate to such an extent ,but if allowed to persist *S. pyogenes* can eventually become the flesh-eating bacteria, devouring the fascia: necrotizing fasciitis.

Drug resistance is the reduction in effectiveness of a drug such as an antimicrobial or an antineoplastic in curing a disease or condition. When the drug is not intended to kill or inhibit a pathogen, then the term is equivalent to dosage failure or drug tolerance. More commonly, the term is used in the context of resistance that pathogens have "acquired", that is, resistance has evolved. When an organism is resistant to more than one drug, it is said to be multidrug-resistant. In a broad sense, the immune system of an organism is such a drug delivery system, albeit autonomous, and faces

the same arms race problems as external drug delivery.

Novel drug targets are required in order to design new defenses against antibiotic-resistant pathogens. Comparative genomics provides new opportunities for finding optimal targets among previously unexplored cellular functions, based on an understanding of related biological processes in bacterial pathogens and their hosts(Kanehisa 1997 and Singh *et al* ,2011) Further target prioritization involves comparison of the corresponding pathways and individual functions between pathogens and the human host. The most promising targets are validated by direct knockouts in model pathogens. The prolonged use of the antibiotics over the years has transformed many organisms resistant to multiple drugs. This has made the field of drug discovery of vital importance in curing various infections and diseases. The drugs act by binding to a specific target protein of prime importance for the cell's survival(Singh *et al* ,2012 and Singh *et al* ,2011).

MATERIALS AND METHODS

The KEGG database can be utilized for modeling and simulation, browsing and retrieval of data. (http://www.genome.ad.jp/kegg/)(Kanehisa *et al*,2004) Proteins were categorized according to their amino acid length. Database of Essential Genes (http://www.essentialgene.org/) was searched with BLAST program (http://blast.ncbi.nlm.nih.gov/Blast) to filter essential genes. Cello is a tool to predict protein subcellular localization using SVM for eukaryotes and Gram negative bacteria. The SWISS-MODEL Workspace is a web-based integrated service dedicated to protein structure homology modeling(Kiefer *et al* ,2009) VMD is designed for modeling, visualization, and analysis of biological systems such as proteins, nucleic acids, lipid bilayer assemblies, etc.

RESULTS AND DISCUSSION

Table 1: Result of KEGG metabolic pathway

S. No	Gene Name	Pathway
1.	SPy_1283	spy00010 (Glycolysis/Gluconeogenesis) spy 00030 (Pentose phosphate pathway)spy00051 (Fructose and mannose metabolism)spy00052 (Galactose metabolism)spy00680 (Methane metabolism)spy01100 (Metabolic pathways) spy01110 (Biosynthesis of secondary metabolites) spy01120 (Microbial metabolism in diverse environments)
2.	SPy_0890	spy00030 (Pentose phosphate pathway)spy00230 (Purine metabolism)
3.	SPy_0399	spy00240 (Pyrimidine metabolism)spy01100 (Metabolic pathways)
4.	SPy_1849	spy00620 (Pyruvate metabolism)spy00640 (Propanoate metabolism)spy00650 (Butanoate metabolism)spy01100 (Metabolic pathways)
5.	SPy_1652	spy00760 (Nicotinate and nicotinamide metabolism)spy 01100 (Metabolic pathways)

The result obtained in the study of structure identification by metabolic pathway of *S. pyogenes* has been given below with their specific genes, though compared and analyzed by using a tool KEGG Pathway (Table 1). The *in silico* method reduces the time as well as the cost of target screening. Over 1400 bacterial genomes have been completely sequenced, providing a substantial amount of data that requires the development of *in silico* techniques for the prediction, identification and selection of targets for investigation .Current antibiotics have often targeted proteins involved in cell maintenance or structure. A case study of the integration of six human metabolic pathways from KEGG depicts the ability of KEGG converter to automatically produce merged and converted to SBML fully functional pathway models (Kanchisa *et al*,2006). Moreover, KEGG converter permits the inclusion of additional reactions in the resulting model which represent flux cross-talk with neighbouring pathways, providing in this way improved simulative accuracy. These additional reactions are introduced by exploiting relevant semantic information for the elements of the KEGG Pathways database. The architecture and functionalities of the web-based application are presented.

Analysis using CELLO, showed that essential *S. pyogenes* strain *M*1 proteins were membrane proteins and thus more amenable as drug targets(Gardy *et al* ,2003) (Table 2). Further, using the KEGG Automatic Annotation Server, we identified six unique metabolic pathways that exist in this bacterium but not in human .The essential proteins that have not been previously characterized were studied using the SVM-Prot.

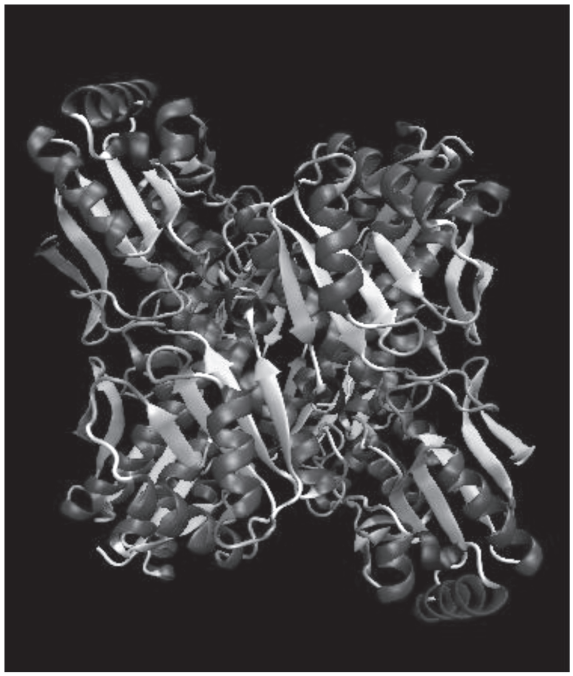

Figure 2: Ribbon model of drug target visualized using VMD Software

Table 2: Result of structure prediction & CELLO

Gene Name	Structure	Organisms	Svm	Localization
SPy_1283	Protein Model Portal	*Leishmania donovani*	Amino Acid CompCELLO Prediction	Cytoplasmic
		Entamoeba histolytica		Cytoplasmic
SPy_0890	Protein Model Portal	*Bacillus cereus*	Amino Acid Comp CELLO Prediction	Cytoplasmic Cytoplasmic
SPy_0399	Protein Model Portal	*S. mitisMycobacterium tuberculosis*	Amino Acid Comp CELLO Prediction	Membrane Cytoplasmic
SPy_1849	HSSP	*Haemophilus influenzaeS. pneumoniae*	Amino Acid Comp CELLO Prediction	Cytoplasmic Cytoplasmic
SPy_1652	Protein Model Portal	*Escherichia coli Mycobacterium tuberculosis*	Amino Acid Comp CELLO Prediction	Cytoplasmic Cytoplasmic
SPy_1233	Protein Model Portal	*Escherichia coli Helicobacter pylori Pseudomonas aeruginosa Mycobacterium tuberculosis*	Amino Acid Comp CELLO Prediction	Cytoplasmic Cytoplasmic

Table 3: Plot Statistics

S.No	Target	Accuracy	Ramachandran Plot	Z- Score
1.	6-phosphofructokinase predicted model	68.32% of the residues had anaveraged 3D-1D score > 0.2 Warning	Residues in most favoured regions [A,B,L] 898 91.3% Residues in additional allowed regions [a,b,l,p] 76 7.7%Residues in generously allowed regions [~a,~b,~l,~p] 10 1.0% Residues in disallowed regions 0 0.0%—— ——— Number of non-glycine and non-proline residues 984 100.0%	0.62

The results obtained in this by using structure validation, where the target was taken from SWISS-MODEL Workspace (Arnold *et a*l ,2006). It has 337 amino acids. To being with, the sequence details have been given below (Table 3).

Visualizing the output data opens the simulator output to human inspection, especially if the output is in the form of 3D graphics that can be assessed qualitatively instead of in tables of textual data (Figure 1). For example, even though the structure of a protein may be described completely in a text file that follows the PDB format, the data take on more meaning when they can be visualized as a 3D structure that can be rotated in 3D space using a visualization program such as VMD.

Thus in the present study (Figure 2) , the 3D model predicted has more than 80% amino acid in the most favored regions and no amino acid in the disallowed region as shown by the Ramachandran Plot.

CONCLUSION

Target-based approaches have provided a valuable inventory of essential genes that can be used to select and validate antimicrobial agents. By comparing the genome of *S. pyogenes* with those of *S.typhi* and *S. aureus*, a total of six target genes were selected and identity with which can be used for drug designing and Structure prediction of homology modeling for 6-phosphofructokinase can be utilized for inhibitor designing.

ACKNOWLEDGEMENT

The authors are grateful to the Sam Higginbottom Institute of Agriculture, Technology and Sciences (Formerly Allahabad Agriculture Institute) (Deemed-to-be-University) Allahabad for providing the facilities and support to complete the present research work.

REFERENCES

F.J. Rosenbach,1884. *Mikro-Organismen bei den Wund-Infections-Krankheiten des Menschen*. D387-389.

Bisno, A.L., Stevens, D.L. 1996. Streptococcal infections of skin and soft tissues. *N. Engl. J. Med*. **334:** 240-245.

Bisno, D.E. 1991. Group A streptococcal infections and acute rheumatic fever. *N. Engl. J. Med*. **325:** 783-793.

Falagas, M.E., Vouloumanou, E.K., Matthaiou, D.K., Kapaskelis, A.M., D.E. Karageorgopoulos, Effectiveness and safety of short-course vs long-course antibiotic therapy for group a beta hemolytic streptococcal tonsillopharyngitis: a meta-analysis of randomized trials. *Mayo Clin Proc*. **83** (2008) 880–889.

Kanehisa, M. 1997. A database for post-genome analysis. *Trends Genet*. **13:**375–376.

Singh, S., Singh, G., Singh, A. K., Gautam, B., Farmer, R., Lodhi, S. S. and Wadhwa, G. 2011. Prediction and Analysis of Paralogous Proteins in *Trichomonas vaginalis* Genome. *Bioinformation* **6(**1): 31-34

Singh, S., Singh, G., Sagar, N., Yadav, P. K., Jain P.A., Gautam, B. and Wadhwa, G. 2012. Insight into *Trichomonas vaginalis* genome evolution through metabolic pathways comparison. *Bioinformation* **8(**4):189-195

Singh, S., Singh, G., Singh, A.K., Gautam, B., Farmer, R., Lodhi, S. S. and Wadhwa, G. 2011. Prediction and Analysis of Paralogous Proteins in *Trichomonas vaginalis* Genome. *Bioinformation* 6(1): 31-34

Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, M. Hattori, 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res*. **32:** D277-280.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L. and Schwede, T. 2009. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research*. **37:**2009 D387-392.

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M. and Kawashima, S. 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*. **34** D354-357.

Gardy, J.L., Spencer, C., Wang, K., Ester, M., Tusnady, G.E., Simon, I., Hua, I., deFays, K., Lambert, C and Nakai, K. 2003. PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria. *Nucleic Acids Res*. **31:**3613-3618.

K. Arnold, L. Bordoli, J. Kopp and Schwede,T. 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. *Bioinformatics*, **22:**195-201.