Advances in Bioresearch
Adv. Biores., Vol 8 (6) November 2017: 102-106

Print ISSN 0976-4585; Online ISSN 2277-1573
Journal's URL:http://www.soeagra.com/abr.html
CODEN: ABRDC3
DOI: 10.15515/abr.0976-4585.8.6.102106

Advances in Bioresearch

ORIGINAL ARTICLE

Comparative Modeling and Functional annotation of Streptococcus pyogenes M1 GAS hypothetical protein SPy _ 0012

S.Sugunakala[1]*,S.Sethupathy[2]
* 1. Department of Bioinformatics, A.V.C.College (Autonomous), Mannampandal, Mayiladuthurai, Tamilnadu, India.
2. Department of Medical Biochemistry, Rajah Muthiah Medical college,Annamalai University, Chidambaram, Tamilnadu, India.
E mail: suguna_murali@hotmail.com

ABSTRACT
In the Drug discovery process, identification of the target protein is an important step. In this study, potential target identification was done on the Streptococcus pyogenesM1GAS which is a human pathogenic organism. The availability of complete genome of S.pyogenes provides a wide opportunity to screen the drug targets. Even if the genome of an organism is completely sequenced, some of the protein sequences are not annotated completely. The proper and complete annotation of proteins is required to understand the pathogenicity and resistance function. The genome of S.pyogenes consists of 1801 genes representing 1693 proteins and many were described as hypothetical proteins. In the present work, the structural and functional annotation was predicted for Streptococcus pyogenes M1 GAS hypothetical protein SPy _ 0012 by using various bioinformatics tools. Analysis for homology and sub cellular localization was done by Blast p and PSORTb tool which showed that this protein is organism specific and not having any homology with the Homo sapiens and it may be localized in the cytoplasmic region . Then, the pathway analysis with KASS tool from KEGG server reveals that it is involved in the betalactam resistance activity. Finally, the primary, secondary structures were analyzed and tertiary structure was predicted by the different tools available in the Expasy server. Validation was done for the predicted 3D structure and it was submitted in the Protein Modeling Data Base and received PMDB ID of PM0080771. Further, putative binding region for the modeled protein was also identified using COACH server. Based on the nature of amino acids present in the ligand binding site, an effective beta lactamase inhibitor could be developed in future to enhance the spectrum of beta lactam antibiotics against S.pyogenes.
Key words: *Streptococcus pyogenes, Structure prediction, comparative modeling, Active site prediction,*

Received 24.05.2017 Revised 25.06.2017 Accepted 29.08.2017

How to cite this article:
S.Sugunakala, S.Sethupathy. Comparative Modeling and Functional annotation of Streptococcus pyogenes M1 GAS hypothetical protein SPy _ 0012. Adv. Biores., Vol 8 [6] November 2017.102-106.

INTRODUCTION

Streptococcus pyogenes also known as Group A Streptococcus (GAS) is a gram positive, facultative, anaerobic microorganism. It has been identified as a leading human pathogenic organism of worldwide morbidity and mortality [1, 2, 3 & 4]. It is associated with a wide variety of disease such as pharyngitis [5], impetigo in children[2], necrotizing fasciitis [6], cellulitis, pneumonia [8] and acute rheumatoid fever [9, 10]etc., Disease caused by *S.pyogenes* are commonly treated with basic antibiotics like penicillin. But treating antibiotic resistance group of pyogenes is still a challenging task in the medical field. Identifying newer targets involved in the resistance property may help to design some novel drug compounds [11]. With the advancement of whole genome sequencing processes, a total of 302 strains of *S.pyogenes* were completely sequenced and all these are available in the NCBI database (URL: www.ncbi.nlm.nih.gov). Among these, GAS SF370[12] is available from the year 2001, and is considered to be a significant strain[13]. At present, it is noted that nearly 10% of protein sequences are not annotated and are denoted as Hypothetical proteins. For the structural and functional characteristics of these proteins, there is no experimental evidence available so far. Prediction of tertiary structure and function for these hypothetical proteins by laboratory based method is a time consuming and expensive method. Use of bioinformatics tools and software provides important structural and functional annotations to this hypothetical protein. Many research works were carried out to identify novel drug target in *S.pyogenes* and in various microorganisms [11, 14 & 15]. Based on the success of these works, the study of the *S.pyogenes M1GAS* hypothetical protein in SF 370 strain was selected. The study was designed to explore its sequence similarity with Homo sapiens proteome databases, essentiality of this protein in the organism, prediction of possible location in the organism, domain, family and functional identification , the structural analysis namely primary, secondary and tertiary structures by using various bioinformatics tools. Finally, possible active sites were also predicted to aid in the development of novel drug molecule to treat pathogenic *S.pyogenes* effectively.

MATERIALS AND METHODS

Sequence retrieval

The whole genome of *Streptococcus pyogenes M1 GAS* consists of 1801genes and 1693proteins. Among these149 were identified as hypothetical protein in which the function and structural annotations are not available. From these hypothetical proteins, the protein sequence of SPy _ 0012 was selected. The amino acid sequence was retrieved from NCBI database (www. ncbi.nlm.nih.gov) with accession number of NP_268432.1.

Similarity analysis of hypothetical protein SPy _ 0012against H. sapiens proteome

Proteins present in the microorganism which are not having any homology to human proteome may serve as an effective drug target. To identify the uniqueness, Blast P analysis was carried out with the retrieved *Streptococcus pyogenes* hypothetical protein SPy _ 0012 against H.sapiens proteome. The parameters like threshold for expectation value of 0.005, and minimum bit score value of 100 were used for filtering the hits.

Sequential Analysis of hypothetical protein SPy _ 0012

Prediction of sub cellular localization may be of help in genome annotation and the understanding functional characteristics[16]. In this study, the PSORTbsub cellular localization prediction tool (http://www.psort.org/psortb/) was used. Using the Database of Essential Genes [17](DEG) (http://tubic.tju.edu.cn/deg/), the *S. pyogenes* hypothetical protein was scanned against 46 bacterial proteomes by bacterial Blast P. Cutoff of $10^{-0.05}$and a minimum bit score of 100 were used as parameters. Further, the conserved domain and family annotation was done with NCBI Conserved Domains Database (NCBI-CDD) [18], and Proteins Families Database (Pfam)[19]. Finally the possible involvement of this hypothetical protein in metabolic pathways was identified by using KEGG automatic annotation server (KAAS) [20].

Structure analysis of hypothetical protein SPy _ 0012

The primary and secondary structure analysis was performed by using Protparam [21] and Sopma tools [22] from Expasy server (http://www.expasy.org/). The tertiary structure for hypothetical protein SPy _ 0012 waspredicted by homology modeling using Protein structure prediction $(ps)^2$ – v2web server [23]. (http://ps2.life.nctu.edu.tw/). This predicted protein structure was aligned with the template structure using Flexible Structure Alignment by Chaining Aligned fragment pairs allowing twists (FATCAT tool) available through the URL: http://fatcat.burnham.org/[24].

Binding site prediction

Ligand binding sites were predicted by using COACH server (http://zhanglab.ccmb. med. umich.edu/COACH/) which uses two different methods namely TM – Site & S- Site. The identified templates will be searched for binding sub specific structure and provides sequence profile comparison [25].

RESULT AND DISCUSSION

If the host proteins are homologous to the drug targeted proteins of the organism, it will produce unwanted therapeutic effects. If the protein present is unique to that organism and is not present in the human being, it could serve as a good drug target [26] and this criterion was fulfilled by this hypothetical protein. Blast P results indicated that this protein did not have any homology with the H. sapiens. Results obtained from the sub cellular localization analysis show that this *S.pyogenes M1 GAS* hypothetical protein SPy_0012 may be localized in cytoplasmic region. This kind of study of sub cellular localization provides some basic knowledge in understanding the mechanism of the proteins involved in disease and may helpin developing new drugs[27]. Further, analysis of predicting essentiality using DEG revealed that it is a non essential protein. The domain analysis by Pfam& CDD databases revealed that this hypothetical protein consists of Betalactamase enzyme family like domain and it is found to be 193 amino acids length which are in the region of 213 – 406. From the KASS server, it is noticed that this hypothetical protein may be involved in the beta - lactam resistance pathway. The primary and secondary structural analysis reveals that this hypothetical protein consists of 428 amino acids, mol. weight of 48848.87, Isoelectric point of 8.14, instability index of 32.93, negative GRAVY index of -0.331. These values suggests that this protein is positively charged, stable, hydrophilic and soluble protein. Presence of higher percentage of leucine amino acid indicates that the regions may favors helices. The results of secondary structure analysis support the view that the *S.pyogenes* hypothetical protein consists of 42.99%, 18.22%, 10.28% and 28.5% of helices, strands, beta turns and coils respectively. The highest percentage contribution of helices favors the protein to fold flexibly and may increase the protein interactions. The process of tertiary structure prediction starts with template identification. Results from template searching process show that the PDB ID: 1iysA produces maximum sequence similarity of 74.5%. Using this PDB structure as a template structure, tertiary structure for *S.pyogenes* hypothetical protein SPy_0012 was modeled (Figure 1). Then this modeled structure was energy minimized by using Gromos 96 force field using SPDB Viewer [28]. The structure alignment of predicted Vs template structure results show that the two structures are significantly similar with p – value of **0.00**$^{e+00}$and have 226 equivalent positions with an RMSD of 1.19 without twists which proves the reliability of the predicted 3D structure (Figure 2). Finally, this model was validatedand deposited in the Protein Model Database (PMDB) and assigned the PMDB ID as PM0080771 [29].

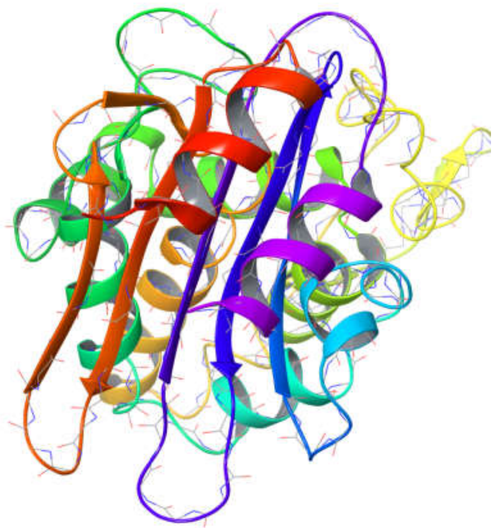

Figure 1: 3D Structure of target protein predicted by using (ps)^{2} – v2 web server

Figure 2: 3D Structure alignment of predicted structure(in grey color) and its template structure PDB ID: 1iysA (in red color).

Binding site prediction

Totally six different binding regions were identified for the modeled hypothetical protein Figure 3. The obtained confidence score, cluster size, PDB hit, ligand name and consensus binding residues of the binding sites are all given in Table 1.

Table 1- Results of binding pocket analysis

Pocket number	C - Score	Cluster Size	PDB Hit	Ligand name	Consensus binding residues
1.	0.95	697	5a93A	PCZ	40,41,44,75,76,109,111,176,192,193,194,195,196
2.	0.05	36	5ee8A	MA4	193,200,225,227,228,229,232
3.	0.04	36	4mbhA	MA4	186,189,204,206,214,216,227,228,231,232
4.	0.01	16	4de3A	DN8	50,54,128,129,132
5.	0.01	9	4a5rB	TBE	41,75,111,142,194,195,196,197
6.	0.01	5	4ddyA	DN6	76,79,109,176,195

Figure 3: Prediction of Binding site

CONCLUSION

In this study, by using Bioinformatics approach, 3D structure for a hypothetical protein SPy_ 0012 of*Streptococcus pyogenes* was modeled and possible functional annotations were also predicted. The diseases caused by *S.pyogenes* were generally treated with beta lactam antibiotics. And the results from functional annotation studies revealed that this hypothetical protein acts as betalactamase and it is involved in the beta lactam resistance activity by producing hydrolytic degradation on the beta lactam ring of antibiotic. In this work, 3D structure for this hypothetical protein was modeled by using 1iys – chain A (Crystal structure of Class A beta Lactamase Toho -1) as template structure. Further, with reference to PDB ID: 5a93 – Chain A (which explores the mechanism of inhibition of beta lactam antibiotic namely Cefotoxim i.e. [(2R)-2-[(1R)-1-{[(2Z)-2-(2-amino-1, 3-thiazol- 4-yl)-2- (methoxyimino) acetyl] amino}-2-oxoethyl]- 5-methylidene-5,6-dihydro-2H-1,3-thiazine- 4-carboxylic acid] by class A Beta lactamase enzyme) [30] with high confidence score (0.95) a putative ligand binding region consists of Gln40,Thr 41,Gly 44,Leu75,Gln76, Ile 109, Lys 111, Ser 176, Tyr 192, Asp 193, Lys 194,Ala195,Phe196was also identified in the modeled hypothetical protein. Based on thenature of amino acids present in the ligand binding site, an effective beta lactamase inhibitor could be developed in future to enhance the spectrum of beta lactam antibiotics against *S.pyogenes*. Further, the same kind of methodology may also be applied to understand the functional and structural aspects of other hypothetical proteins.

REFERENCES

1. Severin, A., Nickbarg, E., Wooters, J., Quazi, S.A., Matsuka, Y.V., Murphy, E., Moutsatsos, I.K., Zagursky, R.J. & Olmsted, S.B. (2007). Proteomic analysis and identification of Streptococcus pyogenes surface-associated proteins. J bacteriol, 189(5): 1514-1522.
2. Todar, K. (2008). Streptococcus pyogenes and streptococcal disease. Online Textbook of Bacteriology.
3. Lamagni, T.L., Darenberg, J., Luca-Harari, B., Siljander, T., Efstratiou, A., Henriques-Normark, B., Vuopio-Varkila, J., Bouvet, A., Creti, R., Ekelund, K. &Koliou, M. (2008). Epidemiology of severe Streptococcus pyogenes disease in Europe. J clinmicrobiol, *46*(7), pp.2359-2367.

4. Ralph, A.P. &Carapetis, J.R. (2012). Group a streptococcal diseases and their global burden. In Host-Pathogen Interactions in Streptococcal Diseases. Springer Berlin Heidelberg, pp. 1-27.
5. Maltezou, H.C., Tsagris, V., Antoniadou, A., Galani, L., Douros, C., Katsarolis, I., Maragos, A., Raftopoulos, V., Biskini, P., Kanellakopoulou, K. and Fretzayas, A. (2008). Evaluation of a rapid antigen detection test in the diagnosis of streptococcal pharyngitis in children and its impact on antibiotic prescription. J antimicrobChemother, *62*(6), pp.1407-1412.
6. Goldberg, G.N., Hansen, R.C. & Lynch, P.J. (1984). Necrotizing fasciitis in infancy: report of three cases and review of the literature. PediatrDermatol, *2*(1): 55-63.
7. Shaunak, s., Wendon, J., Monteil, M & Gordon, A.M. (1988). Septic scarlet fever due to Streptococcus pyogenes cellulitis. QJM, *69*(2): 921-925.
8. Barnham, M., Weightman, N.C., Anderson, A.W., Pagan, F.S. & Chapman, S.T. (1999). Review of 17 cases of pneumonia caused by Streptococcus pyogenes. Euro J ClinMicrobiol Infect Dis, *18*(7): 506-509
9. Bisno, A.L. (1991). Group A streptococcal infections and acute rheumatic fever. N Engl J Med, *325*(11): 783-793.
10. Cunningham, M.W. (2000). Pathogenesis of group A streptococcal infections. ClinMicrobiol Rev, *13*(3): 470-511.
11. Singh, N., Kanojia, H., Singh, S., Verma, D.K., Gautam, B. and Wadhwa, G., 2013. Structure prediction of drug target identified by metabolic pathway analysis of Streptococcus pyogenes. Int J BioinfoBiolSci, 1(1): 79-85.
12. Ferretti, J.J., McShan, W.M., Ajdic, D., Savic, D.J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A.N., Kenton & S. and Lai, H.S. (2001). Complete genome sequence of an M1 strain of Streptococcus pyogenes. Proceedings of the National Academy of Sciences.,98(8): 4658-4663.
13. Okamoto, A. &Yamada, K. (2011). Proteome driven re-evaluation and functional annotation of the Streptococcus pyogenes SF370 genome. BMC microbiology, 11(1).
14. Gupta, D., Banerjee, S., Pailan, S., Saha, P. &Saha–E-mail, P. (2016). In silico identification and characterization of a hypothetical protein of Mycobacterium tuberculosis EAI5 as a potential virulent factor. Bioinformation, 12(3): 182-191.
15. Sharma, P., Patil, K., Sarang, D. &Shinde, P. (2013). In silico structure modeling and characterization of hypothetical proteins present in human fetal brain. Int J AdvBioinforComputBiol, 22
16. Nancy, Y.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J & Brinkman, F.S. (2010). PSORTb 3.0: improved protein sub cellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics.,*26*(13): 1608-1615.
17. Luo, H., Lin, Y., Gao, F., Zhang, C.T. & Zhang, R. (2014). DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. Nuc. Acids Res., *42*(D1): D574-D580.
18. Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R. &Gwadz, M. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. Nuc. Acids Res., *39*(suppl 1): D225-D229.
19. Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K. & Holm, L. (2009). The Pfam protein families database. Nuc. Acids Res., 985.
20. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C. &Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. Nuc. Acids Res., *35*(suppl 2): W182-W185.
21. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.E., Wilkins, M.R., Appel, R.D. &Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. Humana Press pp. 571-607.
22. Geourjon, C. &Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. ComputApplBiosci, 11(6): 681-684.
23. Chen, C.C., Hwang, J.K. & Yang, J.M. (2009). (PS) ^{2}-v2: template-based protein structure prediction server. Bmc Bioinformatics,
24. Ye, Y. &Godzik, A. (2003). Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics, 19(suppl 2): 246-255.
25. Yang, J., Roy, A & Zhang, Y. (2013). Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment, Bioinformatics, 29:2588-2595.
26. Butt, A.M., Batool, M. & Tong, Y. (2011). Homology modeling, comparative genomics and functional annotation of Mycoplasma genitalium hypothetical protein MG_ 237. *Bioinformation*, *7*(6): 299-303.
27. Imai, K., Asakawa, N., Tsuji, T., Akazawa, F., Ino, A., Sonoyama, M. &Mitaku, S. (2008). SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in Gram-negative bacteria. Bioinformation, 2(9): 417-421.
28. Guex, N. &Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. Electrophoresis, *18*(15): 2714-2723.
29. Castrignano, T., De Meo, P.D.O., Cozzetto, D., Talamo, I.G. &Tramontano, A. (2006). The PMDB protein model database. Nuc. Acids Res., 34(suppl 1): D306-D309.
30. Vandavasi, V.G., Weiss, K.L., Cooper, J.B., Erskine, P.T., Tomanicek, S.J., Ostermann, A., Schrader, T.E., Ginell, S.L. & Coates, L. (2015). Exploring the mechanism of β-Lactam Ring Protonation in the Class A β-lactamase Acylation Mechanism using neutron and X-ray crystallography. *J med chem*, *59*(1): 474-479.