IJPSR (2014), Vol. 5, Issue 4 **(Research Article)**

INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES AND RESEARCH

Received on 06 November, 2013; received in revised form, 12 January, 2014; accepted, 26 March, 2014; published 01 April, 2014

IN SILICO CHARACTERIZATION OF GlmU ANTIGENIC *MYCOBACTERIUM TUBERCULOSIS* PROTEIN

Mubashshira A. Pathan*, Aasma S. Jamkhandi and Anil Kumar S Katti

Department of P.G. Studies in Bioinformatics, Walchand Centre for Biotechnology, Walchand College of Arts and Science, Solapur, Maharashtra, India

Keywords:

GlmU, *Mycobacterium tuberculosis, In silico*, Drug target, Protparam, Modeller

Correspondence to Author:

Mubashshira A. Pathan

Department of P.G. Studies in Bioinformatics, Walchand Centre for Biotechnology, Walchand College of Arts and Science, Solapur, Maharashtra, India

E-mail: mubashshirapathan9999@gmail.com

ABSTRACT: Tuberculosis (TB) is a disease caused by a bacterium called *Mycobacterium tuberculosis.* The bacteria usually attack the lungs. Tuberculosis is a global health threat and infecting about one third of the human population. Despite having a variety of anti-tuberculous drugs and availability of effective chemotherapy and Bacille-Calmette –Guerin (BCG) vaccine, tuberculosis remains a leading infectious killer world-wide. This has prompted an urgent need of new drugs and the identification of new drug targets. GlmU is a bifunctional acetyltransferase/uridyltransferase that catalyses the formation of UDP-GlcNAc from GlcN-1-P. UDP-GlcNAc is a substrate for two important biosynthetic pathway: lipopolysaccharide and peptidoglycan synthesis. The *glmU* protein is essential in *Mycobacterium tuberculosis*, being required for optimal bacterial growth. Since inhibition of GlmU affects peptidoglycan synthesis which often results in cell lysis, *M. tuberculosis* GlmU is a potential anti-tuberculosis drug target. The primary protein sequence analysis was done using protparam tool and secondary structure prediction was done using SOPMA. The NTP_transferase domain and IspD domain found in GlmU protein were identified by SMART and its 3D structure prediction was done using modeller. Model quality was analyzed using PROCHECK and showed 91.7% of residues in most favored region in Ramachandran plot. Further the active site residues of bifunctional GlmU proteins were identified from Catalytic Site Atlas (CSA) and catalytic residues were Glu-385 and Arg-19. This *in silico* sequence and structure analysis is helpful for identification and validation of drug target in *Mycobacterium tuberculosis*.

INTRODUCTION: Tuberculosis is one of the most prevalent infections of human beings. *Mycobacterium tuberculosis* (MTB) is a pathogenic bacterium in the genus *Mycobacterium* and the causative agent of most cases of tuberculosis (TB). It was first discovered in 1882 by Robert Koch, M. tuberculosis has an unusual, waxy coating on its cell surface (primarily mycolic acid), which makes the cells impervious to Gram staining [1]. The physiology of *M. tuberculosis* is highly aerobic and requires high levels of oxygen [2]. It does not retain any bacteriological stain due to high lipid content in its wall, and thus is neither Gram-positive nor Gram-negative; hence Ziehl-Neelsen staining or acid-fast staining, is used. *M. tuberculosis* divides

QUICK RESPONSE CODE

DOI: 10.13040/IJPSR.0975-8232.5(4).1493-99

Article can be accessed online on: www.ijpsr.com

DOI link: http://dx.doi.org/10.13040/IJPSR.0975-8232.5(4).1493-99

every 15-20 hours, which is extremely slow compared to other bacteria [3].

In 1995, an outbreak occurred in a rural area in the United States considered to be at low risk for TB [4]. The strain, designated CDC1551 or CSU 93 by the Centers for Disease Control and Prevention (CDC) attracted special attention because of an unusually high rate of transmission as evaluated by skin test conversion. Infection with CDC1551 was not associated with an obvious increase in active TB cases, nor did any patients have extra pulmonary disease. *M. tuberculosis* CDC1551 is not unusually virulent but rather more immunogenic and that it induces a rapid and vigorous cytokine response [5].

As per the WHO Global TB Report 2011, there were an estimated 8.8 million incident cases of TB globally in 2010. Though India is the second-most populous country in the world, India has more new TB cases annually than any other country contributing to a fifth of the global burden of TB. It is estimated that about 40% of Indian population is infected with TB bacillus (http://www.tbcindia. nic.in).

Even though varieties of anti-tuberculous drugs and effective chemotherapy and Bacille-Calmette – Guerin (BCG) vaccine are available, tuberculosis is highly infectious and treatment options are often limited, due to the emergence of multi-drug (MDR) and extreme drug resistant (XDR) TB strains. This has prompted an urgent need of new drugs and the identification of new drug targets.

Bifunctional N-acetylglucosamine-1-phosphate uridyltransferase/ glucose amine -1-phosphate acetyltransferase (GlmU): UDP-N-acetyl-D-glucosamine (UDP-GlcNAc) is an essential precursor of peptidoglycan and the rhamnose-GlcNAc linker region of mycobacterial cell wall. In *Mycobacterium tuberculosis* H37Rv genome, Rv1018c shows strong homology to the GlmU protein involved in the formation of UDP-GlcNAc from other bacteria. GlmU is a bifunctional enzyme that catalyzes two sequential steps in UDP-GlcNAc biosynthesis. Glucosamine-1-phosphate acetyl transferase catalyzes the formation of N-acetylglucosamine-1-phosphate, and N-acetylglucosamine-1-phosphate uridylyltransferase catalyzes the formation of UDP-GlcNAc. The first activity of GlmU, glucosamine-1-phosphate acetyltransferase, catalyzes the formation of N-acetylglucosamine-1-phosphate (GlcNAc-1-P) from glucosamine-1-phosphate using acetyl-CoA as the acetyl donor. The second activity of GlmU, N-acetylglucosamine-1-phosphate uridylyl transferase, catalyzes the formation of UDP-GlcNAc from GlcNAc-1-P and uridine triphosphate (UTP). Hence inhibits peptidoglycan synthesis and results in cell lysis. Thus *M. tuberculosis* GlmU is a potential anti-tuberculosis drug target [6].

METHODS:

Primary sequence analysis: The sequence of bifunctional N-acetylglucosamine-1-phosphate uridyl transferase/glucosamine-1-phosphate acetyltransferase (GlmU) protein from *Mycobacterium tuberculosis* strain CDC1551 having accession number NP_335483.1 was retrieved from NCBI (www.ncbi.nlm.nih.gov/ protein) in FASTA format. The primary sequence, the basic physico-chemical properties, of protein were analyzed using ProtParam (http: //web. expasy .org/protparam/) ProtParam is available through the ExPASy server [7].

Secondary structure analysis: Secondary structure of NP_335483.1 was analysed using SOPMA. It stands for Self Optimized Prediction method available on (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html). SOPMA correctly predicts the secondary structure α-helix, β-sheet and coil [8].

Identification of Domain and Signal peptides: The domains and signal peptides of GlmU were analyzed using SMART (Simple Modular Architecture Research Tool). SMART is an online resource (http://smart.embl.de/) for the identification and annotation of protein domains and analysis of protein domain architecture [9].

Predicting transmembrane topology: Transmembrane helices and topological domains were predicted using TMHMM server (http://www.cbs.dtu.dk/services/TMHMM). This method is based on Hidden Morkov model, TMHMM correctly predicts transmembrane helices it can discriminate between soluble and membrane protein with both specificity and sensitivity [10].

Prediction of DNA binding protein: To predict whether the protein can binds to DNA or not, DNA Binding Protein Prediction server (http://www.enzim.hu/~szia/dnabind.html) was used. This server predicts DNA-binding protein from its structure or sequence [11].

Homology modeling: Three dimensional structure of protein was built using MODELLER. MODELLER is a computer program for homology or comparative protein structure modelling [12].

Active Site Prediction: The active site residues were identified using Catalytic Site Atlas (CSA).It is available at URL-http://www.ebi.ac.uk/thornton-srv/databases/CSA_NEW/. CSA is a database documenting enzyme active sites and catalytic residues in enzymes of 3D structure [13].

Zinc binding site prediction: Accurate prediction of zinc-binding sites in proteins of unknown function provides important clues for the inference of protein function [14]. TEMSP (http://netalign. ustc.edu.cntemsp/index.php.) 3D TEmplate-based Metal Site Prediction was used to predict zinc binding sites based on protein structures.

Protein function prediction: ESG web server was used for automated protein function prediction (http://dragon.bio.purdue.edu/ESG/). ESG out performs conventional PSI-BLAST and the protein function prediction (PFP) algorithm. It is found that the iterative search is effective in capturing multiple-domains in a query protein, enabling accurately predicting several functions which originate from different domains [15].

RESULTS AND DISCUSSION:

Primary protein sequence analysis: The sequence of GlmU (NP_335483.1) contains 513 amino acids with molecular weight 53520.3 Da. The physicochemical properties of the protein by Protparam tool were shown in **Table 1**. The extinction coefficient of GlmU was 24535 and absorption was 0.1%, indicates how much a protein absorbs light at a certain wavelength. The instability index provides an estimate of the stability of protein in a test tube. The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins.

Upgade Akhilesh *et al.* (2012) also used expasy's ProtParam tool to calculate physiochemical parameters of the three Herpes Simplex Viral proteins such as P04290, P06477 and P04486 and similarly characterization of PtpA protein of *Mycobacterium tuberculosis* was done using ProtParam tool by Merly D. P *et al.* (2012).

TABLE 1: PHYSICO-CHEMICAL PROPERTIES OF GlmU BY ProtParam

No. of amino acids	513
Molecular weight	53520.3 Da
Total number of negatively charged residues (Asp + Glu)	52
Total number of positively charged residues (Arg + Lys)	40
Total no. of atoms	7542
Theoretical pI	5.65
Extinction coefficients	24535
Estimated half-life (mammalian reticulocytes, *in vitro*).	30 hours
Instability index (II)	29.17
Aliphatic index	96.55
Grand average of hydropathicity (GRAVY)	0.012

Secondary structure prediction: The secondary structure of the GlmU protein was predicted using SOPMA. The parameters considered were window size as 17 residues and similarity threshold as 8. The details of secondary structure were shown in Table 2 and graphical view from PDBsum is represented in **Fig. 1**.

Mao *et al.* (2005) predicted the secondary structure of human metapneumovirus attachment (G) protein by using SOPMA and Syyada Jafri *et al* (2012) used SOPMA to predict the secondary structure Gp41 envelope glycoprotein of human.

TABLE 2: SECONDARY STRUCTURE PREDICTION OF GlmU BY SOPMA

Secondary structure	No. of residues	Percentage
Alpha helix	142	27.68%
Extended strand	140	27.29%
Beta turn	58	11.31%
Random coil	173	33.72%

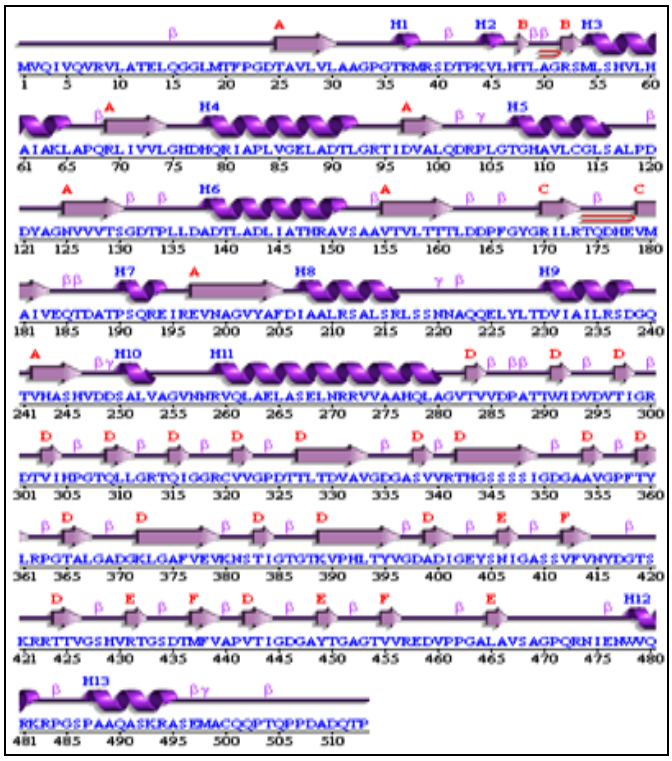

FIG. 1: GRAPHICAL VIEW OF SECONDARY STRUCTURE OF GlmU BY PDBsum

Identification of Domains and Signal peptides: The domains and signal peptides were predicted using SMART. There were two overlapping domains in GlmU protein, NTP_transferase domain and IspD domain. NTP_transferase domain and its sequence were shown in Fig.2, starts at position 26 and ends at position 295 with E-value of 2.00e19. NTP_transferase transfer nucleotides from one compound to another. This domain is found in a number of enzymes that transfer nucleotides onto phosphosugars. IspD domain is shown in Fig.3. It starts at position 25 and ends at position 268 with E-value of 2.00e13. 4-diphosphocytidyl-2C-methyl-D-erythritol synthase, a bacterial IspD protein, catalyzes the third step of the deoxyxylulose-5-phosphate pathway (DXP) of isoprenoid biosynthesis; i.e the formation of 4-diphosphocytidyl-2C-methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-phosphate. The isoprenoid pathway is a well-known target for anti-infective drug development.

Merly D. P *et al* (2012) used SMART and pfam to identify domains in PtpA protein of mycobacterium tuberculosis and domain identified was Tyrosine Phosphatase domains which has catalytic activity. Unlike this Reddy *et al.*, (2009) used PRODOM to identify domain in STK11 protein which was involved in PJ Syndrome.

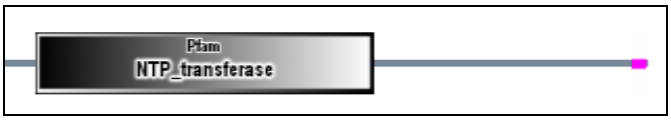

FIG. 2: NTP_TRANSFERASE DOMAIN AND ITS SEQUENCE

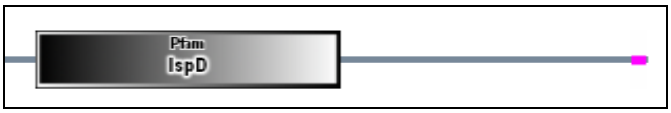

FIG. 3: IspD DOMAIN AND ITS SEQUENCE

Signal peptide: The four signal peptides were predicted using SMART as shown in Fig.4. The sequences of signal peptides along with their position were represented in **Table 3**. The signal peptide is typically between 15 and 40 amino acids long and is essential for protein secretion.

FIG. 4: PREDICTION OF SIGNAL PEPTIDE USING SMART

TABLE 3: SIGNAL PEPTIDES OF GlmU BY SMART

Start position	Sequence	End position
132	DTPLLDADTLADLIAT	147
153	AAVTVLTTTLDD	164
208	AALRSALSRLSSNNA	222
501	QQPTQPPDADQTP	513

Predicting transmembrane topology: Trans-membrane topology of GlmU was predicted using TMHMM server as shown in **Fig. 5**. The protein GlmU was predicted as a secretory protein as it was secreted outside the cell, which means there are no transmembrane helices in the sequence.

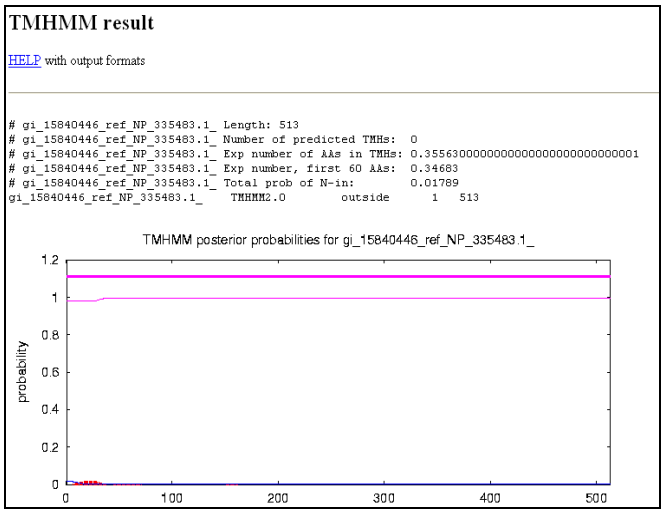

FIG. 5: PREDICTION OF TRANSMEMEBRANE HELICES USING TMHMM SERVER

Prediction of DNA binding protein: The prediction of DNA binding protein was carried out using DNA Binding Protein Prediction server. The threshold value of protein to be considered as DNA binding is 0.362, and predicted value for the sequence is -2.293, hence the protein is predicted as Non DNA binding.

Homology modeling: Homology modeling of GlmU was done using Modeller 9v8. The sequence was searched against PDB for template selection. Four good templates such as 3D8V, 3FOQ, 3SPT and 2QKX were selected with their identities 100%, 100%, 100%, and 99% respectively. Among four PDB templates, 3PST was chosen as best template and four potential models were built by modeller. Finally, the model with lowest DOPE score was selected as shown in **Fig. 6** and further submitted to PROCHECK for model quality analysis.

Ramachandran plot as shown in **Fig. 7** represents the different regions of the plot. Residues in most favored regions are 91.7%, additionally allowed regions 8%. Hence PROCHECK analysis revealed that the model is of good quality. **Table 5** shows the structural evaluation from the Ramachandran plot.

In similar with this Rajasekhar Chikati *et al* (2010) also used modeller 9v8 to generate model of PfSHMT.

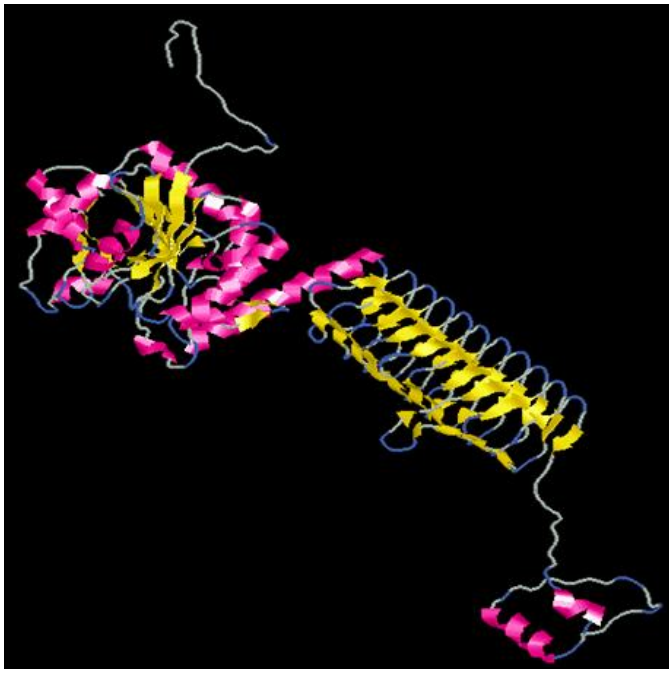

FIG. 6: VISUALIZATION OF GENERATED MODEL BY RASMOL

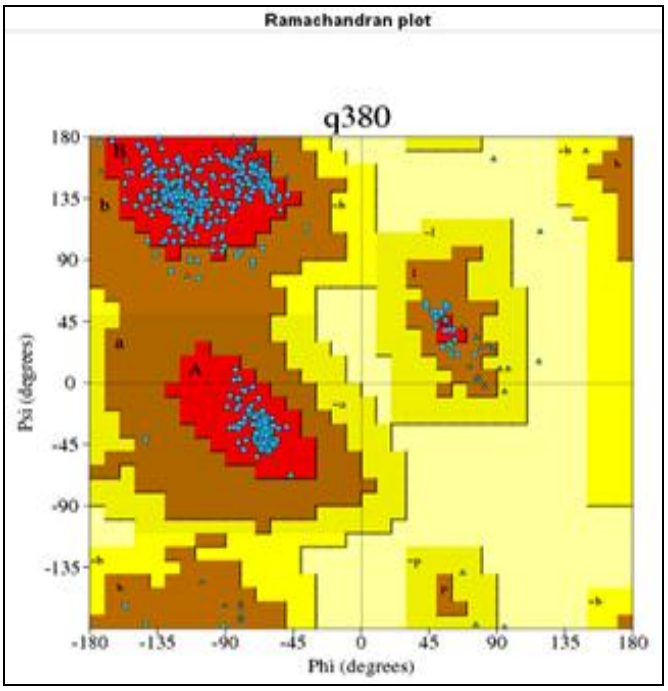

FIG. 7: RAMACHANDRAN PLOT OF GLMU OBTAINED FROM PROCHECK

Active Site Prediction: The active site residues of bifunctional protein GlmU were identified from Catalytic Site Atlas (CSA) and catalytic residues were Glu-385 and Arg-19. The active sites and other structural features of GlmU were visualized using Rasmol as shown in **Fig. 8**.

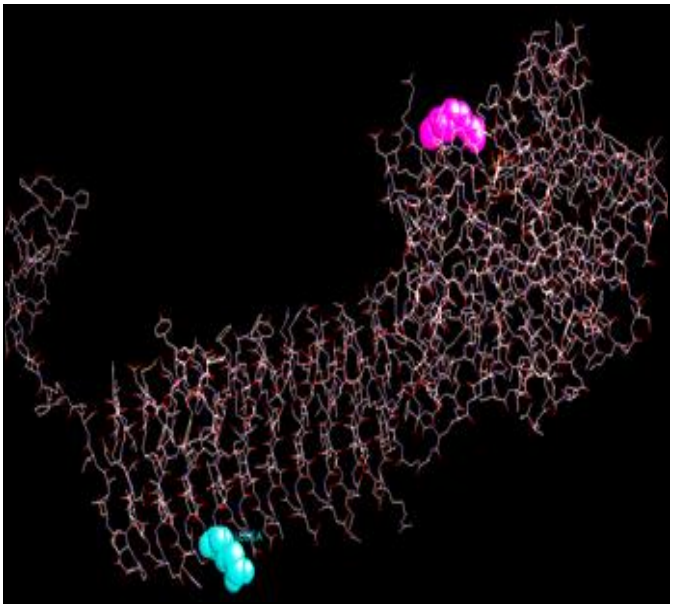

FIG. 8: VISUALIZATION OF ACTIVE SITE RESIDUES IN RasMol (Arg-19 AND Glu-385)

Zinc binding site prediction: TEMSP server predicted GlmU as non-Zinc binding protein. There were no zinc binding residues in the sequence. The zinc binding residues plays very important role in ligand binding.

Protein function prediction: The function of GlmU was predicted using ESG server. ESG reveal the molecular function, biological function and the percentage distribution of the protein in different cellular components as shown in **Table 5, 6 and 7** respectively. ESG predicted that molecular function of GlmU were UDP-N-acetylglucosamine diphosphorylase activity and nucleotidyltransferase activity with 99% probability. GlmU protein was mainly involved in cell wall organization and peptidoglycan biosynthetic process and its subcellular location was cytoplasm.

TABLE 5: MOLECULAR FUNCTION FOR GlmU BY ESG

Probability	Term	Description
99%	GO:0000287	Magnesium ion binding
99%	GO:0003824	Catalytic activity
99%	GO:0003977	UDP-N-acetylglucosamine diphosphorylase activity
99%	GO:0008415	Acyltransferase activity
99%	GO:0016779	Transferase activity
99%	GO:0019134	Nucleotidyltransferase activity

TABLE 6: BIOLOGICAL PROCESS IN WHICH GlmU IS INVOLVED

Probability	Term	Description
99%	GO:0000902	Cell morphogenesis
99%	GO:0007047	Cell wall organization and biogenesis
99%	GO:0008360	Regulation of cell shape
99%	GO:0009103	Lipopolysaccharide biosynthetic process
99%	GO:0009252	Peptidoglycan biosynthetic process
58%	GO:0008299	Isoprenoid biosynthetic process
41%	GO:0009058	Biosynthetic process

TABLE 7: CELLULAR COMPONENT FOR GlmU BY ESG

Probability	Term	Description
99%	GO:0005737	cytoplasm

CONCLUSION: Tuberculosis is life threatening infectious disease caused by *Mycobacterium tuberculosis.* GlmU mycobacterial protein is mainly involved in cell wall synthesis of bacteria and is highly antigenic in human. Its primary sequence analysis was done using Protparam and secondary structure analysis was done using SOPMA. The NTP_transferase domain and IspD domain were found in GlmU protein using SMART. Transmembrane topology predicted by TMHMM server reveals that GlmU protein is secreted protein.

GlmU protein was predicted as non DNA binding protein using DNA Binding Protein Prediction server. The molecular modeling of GlmU protein was carried out using MODLLER. Model quality was analyzed using PROCHECK and showed 91.7% of residues in most favored region in Ramachandran plot. Hence it is a good quality model. The active site residues were Glu-385 and Arg-19.

TEMSP server predicted GlmU as non-Zinc binding protein and function prediction was carried out using ESG server. Molecular functions of GlmU were UDP-N-acetyl glucosamine diphosphorylase activity and nucleotidyl transferase activity with 99% probability. Further this *in silico* analysis should be validated using *In vitro* assays.

REFERENCES:

1. Michael SG, William R J: Microbial pathogenesis of Mycobacterium tuberculosis Dawn of a discipline. Cell 2001, 104:477-485.
2. Ismael K, Ray CG: *Sherris* Medical Microbiology 2004 (4th ed.). McGraw Hill. ISBN 0-8385-8529-9.
3. Murray PR, Rosenthal KS, Pfaller MA: Medical Microbiology 2005, Elsevier Mosby.
4. Valway SE, Sanchez MP, Shinnick TF, Orme I, Agerton T, Hoy D, Jones JS, Westmoreland H, Onorato IM: An outbreak involving extensive transmission of virulent strain of Mycobacterium tuberculosis. N. Engl. J. Med 1998, 338:633.
5. Musser H, Freedman, Bergtold GK, Freeman S, Patrick AJ, Haslett, James M, Claudia Manca, Tsenova L, Clifton E, Barry A: Mycobacterium tuberculosis CDC1551 Induces a More Vigorous Host Response In Vivo and In Vitro, But Is Not More Virulent Than Other Clinical Isolates. J Immunol 1999, 162:6740-6746
6. Wenli Z, Victoria CJ, Michael SS, Sebabrata M, Dean C, Suresh B, Yi Xin, Michael RN, Yufang M: Expression, essentiality, and a microtiter plate assay for mycobacterial GlmU, the bifunctional glucosamine-1-phosphate acetyltransferase and N-acetylglucosamine-1-phosphate

uridyltransferase. Int J Biochem Cell Biol. 2008, 40: 2560–2571.
7. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel R and Bairoch A: ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 2003, (31): 3784–3788.
8. Geourjon C and Deléage G: SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci.* 1995 11(6): 681-684.
9. Ivica L, Tobias D and Peer B: SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research* 2012 (40).
10. Anders k, Bjorn L, Gunner H and Erik L: Predicting Transmembrane Protein Topology with a Hidden markov model Application to Complete Genomes. *J. Mol.Bio* 2001 (305): 567-580.
11. Szilagyi A and Skolnick J: Efficient prediction of nucleic acid binding function from low resolution protein structures. *J Mol Biol.* 2006 358(3): 922-33.
12. Sali A and Blundell T: Comparative protein modeling by satisfaction of spatial restr aints *J Mol.Bio.* 1993 (234): 1753-1773.
13. Craig P, Gail B and Janet T: The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucl. Acids. Res*. 2004 (32): D129-D133.
14. Wei Z, Meng Z, Liang D, Liwen N, Haiyan L And Maikun T: Structure-based de novo prediction of zinc-binding sites in proteins of unknown function. *Structural bioinformatics* 2011 (27)9: 1262–1268.
15. Meghana C, Troy H, Changsoon P and Daisuke K: ESG: extended similarity group method for automated protein function prediction. Bioinformatics 2009 (25)14: 1739–1745.
16. Akhilesh U, Bhaskar A, Issar S, Senthamarai SV: In Silico Characteriza tion of Keratitis Causing Herpes Simplex Virus (HSV 1) Membrane Proteins using Computational Tools and Servers. Research Journal of Recent Sciences 2012, 1:27-31.
17. Merly DP, Anil Kumar S, Reddy KM, Tejaswini B, Darley SP, Sulochana MB: In-silico analysis of PtpA - an antigenic protein of Mycobacterium tuberculosis. Recent Research in Science and Technology 2012, 4(7): 23-26.
18. Mao HW, Zhao XD, Yang XQ: Prediction of the secondary structure and B cell epitopes of human metapneumovirus attachment protein. Chinese Journal Of Microbiology And Immunology 2005, 25: 1031-1034
19. Syyada SJ, Saliha K, Syed BJ, Masaud S: Structure based sequence analysis & epitope prediction of gp41 HIV1 envelope glycoprotein isolated in Pakistan. Genet Vaccines Ther. 2012, 23(10): 10.
20. Reddy MV, Bhanu BD, Prasad M, Ramya N, Madhulika K: In Silico characterization of STK11- a protein involved in PJ Syndrome. *The Internet Journal of Genomics and Proteomics* 2009, 4:1.5580.
21. Chikati R, Anuradha CM, Chaitanya M, Babajan B, Prasad Y, Kumar S: Structural characterization of pfSerine hydroxymethyltransferase: A novel target for malaria. Journal of Bioinformatics and Sequence Analysis 2010, 2:13-24.

How to cite this article:
Pathan MA, Jamkhandi AS and Katti AKS: *In silico* characterization of GlmU antigenic *Mycobacterium tuberculosis* protein. *Int J Pharm Sci Res* 2014; 5(4): 1493-99.doi: 10.13040/IJPSR.0975-8232.5(4).1493-99