Original Article

In silico Study of Target Proteins for *Mycobacterium tuberculosis*

C. B. Bhagat*[1], S. K. Tank[1], P. R. Dudhagara[2], N. D. Trivedi[3] and U. N. Trivedi[4]

[1]Department of Biosciences, Veer Narmad South Gujarat University, Surat, Gujarat, India
[2]Department of Biotechnology, Veer Narmad South Gujarat University, Surat, Gujarat, India
[3]A. R. College of Pharmacy & G. H. Patel Institute of Pharmacy, Vallabh Vidyanagar, Gujarat, India
[4]Shivam Pharmaceutical Studies and Research Center, Valasan, Anand, Gujarat, India

Address for Correspondence
Department of Biosciences, Veer Narmad South Gujarat University, Surat, Gujarat-395007, India.

E-mail: cbb.chintan@gmail.com

ABSTRACT

The completion of the genome of pathogens and the human has provided data that can be utilized to design vaccines and drug targets. One of the recently adopted strategies for drug designing is based on comparative genomics approach, it gives a set of genes that are likely to be essential to the pathogen but absent in the host. By performing homology searches and structural modeling, we can determine which of these proteins can provide novel targets for designing functional inhibitor compounds active against bacteria. In this study, we used three proteins that are potential target, (NCBI Accession no.) NP_216679, NP_218309, NP_218312, for *Mycobacterium tuberculosis*. Physico-chemical characterization, prediction of secondary structure, disulfide bridges and functional characterization was done to interpret their properties. Three dimensional structure of these proteins were not available as yet at PDB. Therefore, homology models for these proteins were developed using Swiss model server, Modeller and ESyPred3D Web Server. The models were validated using PROCHECK. The study proved that NP_218312 was most stable structure and has good stereo chemical properties. Thus it is selected for further Docking study with antimicrobial peptide, aurein 1.2 and with available drug such as Pyrazinamide, Ethambutol and Isoniazid using HEX. Among these Aurein1.2 was found more effective than other drug, indicates this can be used for further drug designing for *Mycobacterium tuberculosis*.

Keywords: Modeling, Docking, *Mycobacterium tuberculosis*, Swiss model, Procheck, Hex, Antimicrobial peptide.

INTRODUCTION

Completion of human and pathogenic bacteria genome provided lots of raw material for *in silico* analysis and drug designing[1]. Identification of essential gene from the genome of bacteria play significant role and are thus important for the survival of an organism and are non-homologous to human genes. Thus, Identification of essential genes of bacteria is one of the promising means to identify novel drug targets. Availability of the genome sequence of pathogens has provided a tremendous amount of information that can be useful in drug target and vaccine target identification. Thus essential genes provide selectivity and yield a drug which is highly selective against pathogen with respect to human host. A subtractive genome approach and bioinformatics provide opportunities for finding the optimal drug targets from proteome analysis[2]. Proteins that cooperate towards a common biological function are located in the sub cellular compartment. Eukaryotic cell has evolved highly elaborated subcellular compartment but prokaryotes (gram negative bacteria) have five major subcellular localizations (outer membrane, the inner membrane, periplasm, cytoplasm and extracellular), specialized in different biochemical process[3,4]. This identification of various subcellular proteins helps in development of drug candidates. In this study three different proteins were selected for a task, they are NP_216679, NP_218309, and NP_218312 [NCBI]. By subjecting these non-structural membrane proteins to homology searches and structural modeling, we can determine which of these proteins can function as most effective surface epitope for the drug. Screening against such novel targets results in designing of functional inhibitors and lead to discovery of novel therapeutic compounds active against bacteria including the increased number of antibiotic resistant clinical strains[5]. Various computational tools available to researcher to explore and understand physicochemical and structural properties of proteins and also to predict lead protein. Due to the drawbacks of experimental methods that have been used to characterize the proteins of various organisms are time consuming, costly and the fact that these methods not amendable to high throughput techniques so *in silico* approaches provide a viable solution to these problems[6,7]. The amino acid sequence provides most of the information required for determining and characterizing molecule's functional, physical and chemical properties. In this study, the *in silico* analysis and homology modeling studies of all predicted target proteins were reported and best was selected for docking study. Three dimensional structure of these proteins has been yet not available. Hence to describe its structural features and to understand molecular function, the model structures for this protein were constructed. Due to antibiotic resistance, There is a necessitate to develop a new class of antibiotics. Natural antimicrobial peptides can serve as critical defense molecules protecting the host from the invasion of bacteria. Although, they have some drawbacks. Nevertheless, antimicrobial peptides can be new hope in developing novel, effective and safe therapeutics without antibiotic resistance. Future more, AMP-mimetics can be useful in better understanding the role and function of host defense peptides and might also considerably increases the potential of uncovering low-cost highly selective derivatives that will be beneficial utilized in various antimicrobial fields, including in the systemic treatment of medical conditions associated with MDR pathogens. The diversity and broad spectrum antimicrobial

activity of AMPs along with its multidimensional properties that could be exploited for the application of these bioactive peptides as a potential and promising drug candidate in pharmaceutical industries. Thus, it is necessary to discover new antimicrobial sources in nature and study their structures and physicochemical properties more in depth[8-10]. So it needs to screen AMP on *In Silico* platform to choose best one for experimental study.

MATERIALS AND METHODS

Sequences of target proteins of *Mycobacterium tuberculosis* were retrieved from the NCBI, a public domain database. The protein sequences were retrieved in FASTA format and used for further analysis.

Physicochemical characterization

For physical-chemical characterization includes the theoretical isoelectric point (pI), Molecular weight (M. wt), Total number of positive and negative residues ($R^{+/-}$), instability index (II)[11], and grand average hydropathy (GRAVY)[12] were committed using expasy's protparam Server[13].

Functional characterization

The SOSUI server performed the identify transmembrane regions[14]. Disulphide bonds are important in determining functional linkages. The prediction of "SS" bonds using primary structure (protein sequence data) by tool CYS_REC (http://sunl.softberry.com/berry.phtmal? topic). CYS_REC identifies the position of cysteine, the total number of cysteine present and patterns, if present, of pairs in the protein sequence[15].

Secondary structure prediction

SOPMA was employed for calculating the secondary structural features of the protein sequences considered for this study[16]. The secondary structure was predicted by using default parameters.

Model building and evaluation

The modeling of 3D structures of the proteins was performed by three homology program, Swiss model and modeller and ESyPred3D Web Server[17-19]. The constructed 3D models are energy minimized using GROMOS. The overall stereochemical property of protein was assessed by Ramchandran plot analysis using procheck[20,21]. The template for protein was selected using BLASTP[22].

Docking

Docking of the best protein with tradition drug such as Pyrazinamide, Ethambutol and Isoniazid and antimicrobial protein, Aurein 1.2, was carried out using Hex[23].

RESULT

Three proteins of *Mycobacterium tuberculosis* were selected in the study (Table 1). Different parameters computed using Expasy's protparam tool. It shows that all protein has an isoelectric point (pI) above 7 showing at all are basic in nature. The protein instability index was found in a narrow range for all proteins, however instability index values for proteins were found to be ranging from 34.48 to 43.17. NP_216679 is unstable and NP_218309 were more stable than NP_218312. GRAVY indices of proteins ranged from -0.346 to 0.438 (Table 2). The server SOSUI classifies all proteins as soluble proteins with varied number of transmembrane region (Table 3). For thermostability study CYS_REC Results show that NP_216679 has three cysteine residues whereas NP_218312 contains eight cysteine residues suggest the greater thermostability (Table 4). Secondary

structure prediction by SOPMA revels that all protein have different secondary structures. The results revealed that random coil dominated among secondary structure elements for NP_216679 and NP_218312 (Table 5). Modeling results by different software indicate that out of three protein sequences, 3D structures were not modeled for NP_218309 by Swiss Model and ESyPred3D Server. The remaining two proteins NP_216679 and NP_218312 for were modeled successfully in 3D structure and their detail analysis was carried out using Ramchandran plot for stereochemistry properties (Table 6). The NP_218312 was selected for docking studies and its stereochemistry was measured (Figure 1). The phi and psi distribution of the Ramchandran Map generated by of non-glycine, non proline residues were summarized in Table 6. The Docking process was performed for four Compound, Pyrazinamide, Ethambutol, Isoniazid and Aurein 1.2 using default parameter by HEX with protein NP_218312. In these, the E_{min} value of Aurein 1.2 [-434.7] is more than another three traditional drugs (Table 7).

DISCUSSION

The Isoelectric point (pI) will be useful as it determines the solubility and mobility of proteins. Computed pI values of all protein were greater than 7 indicate that these proteins were considered as basic. The computed isoelectric point will be useful for developing a buffer system for purification by isoelectric focusing method and also to conclude half life of protein[11,24] . There are certain dipeptides, the occurrence of which is significantly different in unstable proteins compared with those in stable ones and is termed as an instability index (II). The protein whose instability index is smaller than 40 predicted as stable, and those having more than 40 that are considered as unstable. The selected protein, NP_218312, assigns the value 37.13 proving it as stable protein[11]. The grand average hydropathy (GRAVY) value for a peptide or protein is calculated as the sum of hydropathy values of all amino acids, divide by the number of residues in the sequence. The lower range indicates the possibility of better interaction with water. These results indicate that NP_218312 could be a good choice for drug designing as it has lower value[6,12].

Functional analysis of this protein includes prediction of transmembrane region and disulphide bond. SOSUI distinguishes between membrane and soluble proteins from amino acid sequences and predicts the transmembrane helices for soluble proteins. NP_218312 has higher number of transmembrane region among all selected proteins that important factor to be efficacy of drug and also disulphide bridges play important role in determining thermo-stability of these proteins. CYS_REC was used to determine the cysteine residues and disulphide bonds. There is a presence of SS bond in protein NP_218312 and NP_216679 but not in NP_218309. Even, highest number of SS bound found in protein NP_218312 than other all proves its stability[15].

The secondary structures of proteins were predicted by Self Optimized Prediction Method with Alignment (SOPMA) which correctly predicts 69.5% of amino acids for a state description of the secondary structure prediction .The secondary structure indicates whether a given amino acid lies in a helix, strand or coil. The selected protein NP_218312 has most abudance from as coil form.

The modeling of the three dimensional structure of the protein was performed by three homology modeling programs, Swiss Model server and Modeller, EsyPred3D web Server to compare and to choose best model as did by

previous researcher to be accurate[6,7,17-19]. Among all modeled protein NP_218312 found to be more acceptable. The E_{min} value of various drugs computed using docking software hex[7]. Result suggest that Aurein 1.2 has lowest E-value which indicate it's use of antimicrobial drug designing and further work.

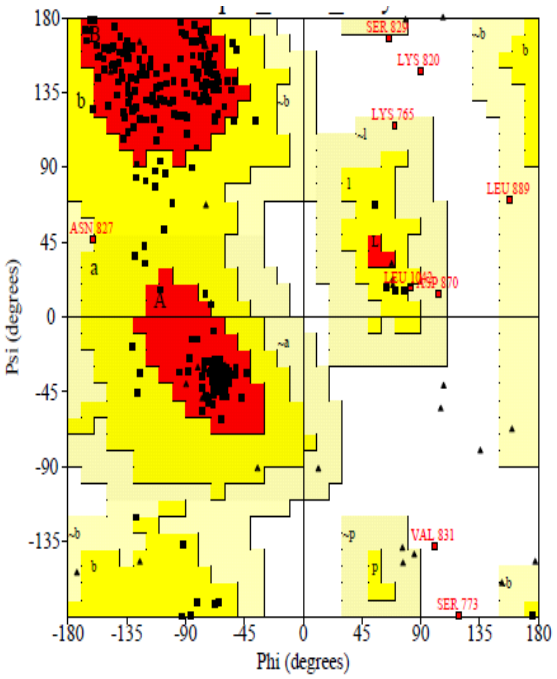
CONCLUSION

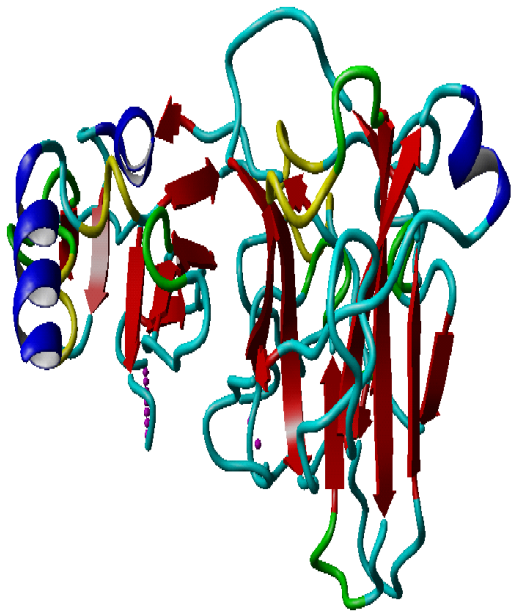
Results indicate NP_218312 was most stable and has good stereochemical properties which make the protein of choice for drug development. Docking result's proved protein as more effective than other drug and could be used for further in vitro study and drug designing. There should be more AMP are to be screened for designing novel antibiotic or bioactive compound against *Mycobacterium tuberculosis*. Thus, this type of study should be carried out on large scale to match best protein and ligand for *in-vitro* study.

REFERENCES

1. Dutta A, Singh SK, Ghosh P, Mukherjee R, Mitter S, Bandyopadhyay D. *In silico* Identification of Potential Therapeutic Targets in the Human Pathogen *Helicobacter Pylori*. *In Silico Biol* 2006; 6: 43-47.
2. Reddy EH, Satpathy GR. Identification of potential targets and lead molecules for designing inhibitory drugs against *Chlamydophila pneumonia*. *Onl J Bioinform* 2009; 10: 14-28.
3. Lu Z, Szafron D, Greiner R, Lu P, Wishart DS, Poulin B, Anvik J, Macdonell C, Eisner R. Predicting subcellular localization of proteins using machine-learned classifiers. *Bioinformatics* 2004; 20: 547-556.
4. Garg A, Raghava GP. ESLpred2: improved method for predicting subcellular localization of eukaryotic proteins. *BMC Bioinf* 2008; 9: 503.
5. Zhang R, Lin Y. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res* 2009; 37: D455-D458.
6. Prajapati C, Bhagat C. In-Silico Analysis and Homology Modeling Of Target Proteins for *Clostridium Botulinum*. *Int J Pharm Sci Drug Res* 2012; 3: 2050-2056.
7. Vaseeharan B, Valli SJ. *In silico* homology modeling of prophenoloxidase activating factor Serine Proteinase gene from the haemocytes of *Fenneropenaeus indicus*. *J Proteomics Bioinform* 2011; 4: 053-057.
8. Rotem S, Mor A. Antimicrobial peptide mimics for improved therapeutic properties. *Biochim Biophys Acta Biomembr* 2009; 1788: 1582-1592.
9. Kanj SJ, Kim DH, Mishig-Ochir T, Lee BJ. Antimicrobial peptides: their physico-chemical properties and therapeutic application. *Arch Pharm Res* 2012; 35: 409-413.
10. Pushpanathan M, Gunasekaran P, Rajendhran J. Antimicrobial Peptides: Versatile Biological Properties. *Int J Pept* 2013; 2013: Article ID 675391.
11. Guruprasad K, Reddy BB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Eng* 1990; 4: 155-161.
12. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982; 157: 105-132.
13. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. The Proteomics Protocols Handbook, edited by JM Walker. Totowa: Humana Press. 2005.
14. Hirokawa T, Boon-Chieng, S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 1998; 14: 378-379.
15. Ferrè F, Clote P. Disulfide connectivity prediction using secondary structure information and diresidue frequencies. *Bioinformatics* 2005; 21: 2336-2346.
16. Geourjon C, Deleage G. SOPM: a self-optimized method for protein secondary structure prediction. *Protein Eng* 1994; 7: 157-164.

17. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 2006; 22: 195-201.
18. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *Protein Struct Distance Anal* 1994; 64: C86.
19. Lambert C, Léonard N, De Bolle X, Depiereux E. ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics* 2002; 18: 1250-1256.
20. Ramachandran GN, Ramakrishnan CT, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol* 1963; 7: 95-99.
21. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 1996; 8: 477-486.
22. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; 215: 403-410.
23. Ritchie DW, Venkatraman V. Ultra-fast FFT protein docking on graphics processors. *Bioinformatics* 2010; 26: 2398-2405.
24. Fred JD, Alfred LD. Relationship between *in vivo* degradative rates and isoelectric points of proteins. *Proc. Nat. Acad. Sci. USA* 1975; 72:3893-3897.





(A) (B)

Figure 1. (A) Ramchandran plot of NP_218312 (B) modeled protein NP_218312

Table 1. Protein sequences included in the study

Sr No.	NCBI Accession number	Length (Amino acid)	Description
1	NP_216679	679	Involved in peptidoglycan biosynthesis
2	NP_218309	643	Act as transmembrane protein
3	NP_218312	1098	Involved in the biosynthesis of the mycobacterial cell wall arabinan

Table 2. Parameters computed using expasy's protparam tool

Accession number [NCBI]	M.W	PI	-R	+R	Instability Index	GRAVY
NP_216679	72537	9.64	60	75	43.17	-0.346
NP_218309	69515	9.85	37	55	34.48	0.438
NP_218312	118021	9.58	70	91	37.13	0.304

Table 3. Transmembrane regions identified by SOSUI server

Accession number	No of transmembrane region	Length	Type of protein
NP_216679	1	23	Membrane
NP_218309	13	23	Membrane
NP_218312	13	23	Membrane

Table 4. Disulphide bond pattern determined by CYS_REC

Accession number	NO OF CYSTEINES FOUND	No. of SS BOUND
NP_216679	3	1
NP_218309	3	None
NP_218312	8	1

Table 5. Calculated secondary structure elements by SOPMA

Protein	NP_216679	NP_218309	NP_218312
Alpha helix	26.22 %	49.77 %	35.61 %
Beta 10 helix	0.0 %	0.0 %	0.0 %
Pi helix	0.0 %	0.0 %	0.0 %
Beta bridge	0.0 %	0.0 %	0.0 %
Extended strand	15.17 %	14.62 %	12.84 %
Beta turn	6.92 %	3.58 %	3.92 %
Bend region	0.0 %	0.0 %	0.0 %
Random coil	51.69 %	32.04 %	47.63 %
Ambiguous state	0.0 %	0.0 %	0.0 %
Other state	0.0 %	0.0 %	0.0 %

Table 6. Ramachandran plot calculation and Comparative analysis of the models from Swiss-model and Modeller computed with the PROCHECK program

Server	Protein	NP_216679	NP_218309	NP_218312
Swiss model	Residues in the most Favored Region	86.7 %	Not Modeled	77.3 %
	Residues in additionally allowed region	10.6 %		17.2 %
	Residues in generously allowed region	2.3 %		3.1 %
	Residues in disallowed region	0.5 %		2.3 %
Modeler	Residues in the most Favored Region	75.7%	80.1%	76.5%
	Residues in additionally allowed region	18.5%	14.8%	16.2%
	Residues in generously allowed region	4.1%	3.3%	4.4%
	Residues in disallowed region	1.8%	1.8%	2.9%
ESyPred3D Web Server	Residues in the most Favored Region	91.0 %	Not Modeled	83.8 %
	Residues in additionally allowed region	7.6 %		12.2 %
	Residues in generously allowed region	1.4 %		2.2 %
	Residues in disallowed region	0.0 %		1.7 %

Table 7. The E_{min} value of various drugs computed using docking process by hex tool

Sr. No.	Drug	E_{min}
1	Pyrazinamide	-125.0
2	Ethambutol	-181.0
3	Isoniazide	-140.5
4	Aurein 1.2	-434.7