ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES
Knowledge to Innovation

Vol 13, Issue 11, 2020

Online - 2455-3891
Print - 0974-2441

Research Article

CONSTRUCTION OF COMPUTATIONAL 3D STRUCTURES OF PROTEIN DRUG TARGETS OF *MYCOBACTERIUM TUBERCULOSIS*

SUDHIR RAO, MISBAH SAYEEDA, TEJASHREE PRAKASH, PADMASHREE AP, SABIA IMRAN, LOKESH RAVI*

Department of Botany, St Joseph's College (Autonomous), Bengaluru, Karnataka, India. Email: lokesh.ravi@sjc.ac.in

Received: 27 July 2020, Revised and Accepted: 26 August 2020

ABSTRACT

Objective: This study aims in constructing a three-dimensional modeled protein structure of potential drug targets in *Mycobacterium tuberculosis* bacteria.

Methods: The protein models were constructed using SWISS-Model online tool. The constructed protein models were submitted in online database called Protein Model Database (PMDB) for public access to the structures.

Results: A total of 100 protein sequences of *M. tuberculosis* were retrieved from UniProt database and were subjected for sequence similarity search and homology model construction. The constructed models were subjected for Ramachandran plot analysis to validate the quality of the structures. A total of 69 structures were considered to be of significant quality and were submitted to the online database PMDB.

Conclusion: These predicted structures would help greatly in identification and drug design to various strains of *M. tuberculosis* that are sensitive and resistant to different antibiotics. This would greatly help in drug development and personalized drug treatment against different strains of the pathogen. This database would significantly support the structure-based computational drug design applications toward personalized medicine in regard to differences in the various strains of the pathogen.

Keywords: *Mycobacterium tuberculosis*, Protein model database, SWISS-Model, Homology modeling, Ramachandran plot.

 DOI: http://dx.doi.org/10.22159/ajpcr.2020.v13i11.39242

INTRODUCTION

Pharmaceutical industries are greatly depend on structure-based computer-aided drug design (SCADD) for manufacturing drugs to treat diseases. It uses three-dimensional (3D) structures of proteins and protein models, to design a drug specific to the target protein. The 3D structures of proteins are usually constructed by analytical techniques such as X-ray crystallography and nuclear magnetic resonance. However, these techniques were too costly to sustain and are time consuming. To tackle these problems, homology modeling aims to develop 3D models of proteins based on similarities in other protein sequences for which crystallographic structure is available, belonging to a different organism. This concept of homology modeling was employed in this study, using Swiss-Model tool to construct 3D structures of known drug target proteins [1-3]. This process of homology model uses computational algorithms to compare, match, analytically predict the 3D coordinates of amino acid sequences, based on pre-existing protein structures that share a significant similarity at sequence level [2,4-6].

Mycobacterium tuberculosis causes a disease called tuberculosis [7-10]. A recent surge in cases involving strains of this bacterium that is resistant to existing antibiotic treatments demands further search for new drugs to combat the disease [9-11]. Tuberculosis claims over 1.3 million lives a year and many survivors continue to suffer through residual organ damage [12]. An established fact is that nearly a third of the world population are asymptomatic carriers of this bacterium [12]. Epidemiologically, this disease is not an endemic disease, however, there is a greater incidence of the disease in tropical and subtropical regions such as Africa and parts of Asia. This disease is primarily of the respiratory system of humans. Complications of other organ systems are also very likely in rare cases of the disease [13]. The first strains of this bacterium were known to affect cattle [14], however, in early 16th century, it was found that the same bacterium had switched hosts and had begun to affect humans with the same disease [8].

The bacterium is classified under the family of Mycobacteriaceae and the fourth international spoligotyping database has described over 39,000 different strains of disease-causing bacteria [11,15]. However, diagnosis of the disease is reliant on primitive acid-fast staining procedures [16]. As a possible application of our database of 3D modeled proteins, there is a scope for the use of 3D modeled proteins to accurately diagnose the strain causing disease in a particular patient, enabling a personalized drug administration [17]. The method has shown promising outcomes in the diagnosis of certain inheritable diseases by the study of the effect of point mutations in 3D modeled proteins. It is found that this bacterium expresses pathogenicity primarily though proteins, which with an accurate database of these proteins can speed up diagnosis [14]. Another application is that our protein model database (PMDB) aids further research and drug development through SCADD.

METHODS

NCBI bibliographic database

The current status of research in the field of protein modeling particularly regarding this species of *M. tuberculosis* was analyzed using NCBI database. The NCBI database (https://www.ncbi.nlm.nih.gov) is a tool that offers a brief understanding of literature and scientific experimentation. This lets us ascertain an appropriate procedure for this study. A search query of the bacterium yielded all that needed to know [1,3,18].

Sequence retrieval

The UniProt Knowledgebase also known as the Universal Protein Knowledgebase (https://www.uniprot.org) is a database that contains non-redundant protein sequences for many organisms. Protein sequences of the various proteins of the bacterium *M. tuberculosis* were

retrieved from this database. It contains amino acid sequences of the protein which are available as a .fasta file and it is stored for further study. The search query in the database consisted of the name of the bacterium without additional parameters [1,2,19].

Sequence alignment

Online server based tool called pBLAST or Protein Basic Local Alignment Search Tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) was used for analyzing the retrieved sequences by comparing it to a pre-existing repository of amino acid sequence data in the Protein Data Bank. This gives results in the form a percentage similarity. The percent match was recorded. However, those proteins with a low percent match were discarded from the study [1,20].

Structure prediction

The amino acid sequence data from the UniProt website in the .fasta file format were subjected to protein modeling. This process constructs the 3D models of the proteins based on the sequence data and an algorithm that compares it to other pre-existing modeled proteins through the concept of homology modeling. For this SWISS-Model (https://swissmodel.expasy.org) was used. The process generates multiple models for each submitted protein which are stored as .pdb file [2,4-6,21].

Model analysis

The SWISS-Model also offers a tool called MolProbity used for qualitative analysis of the protein structure [1]. Ramachandran

Table 1: List of modeled proteins with their PMDB ID

PMDB ID	Protein name	Confidence score (%)	PMDB ID	Protein name	Confidence score (%)
PM0082936	Putative peptide synthase	95.03	PM0082955	Phenolpthiocerol synthesis type-I polyketide synthase PpsC	92.31
PM0082938	Polyketide synthase Pks6	94.51	PM0082956	Phenolpthiocerol synthesis type-I polyketide synthase PpsC	92.31
PM0082939	PPE family protein	100	PM0082957	Phenolpthiocerol synthesis type-I polyketide synthase PpsC	92.31
PM0082940	Carrier domain-containing protein	93.77	PM0082958	Phenolpthiocerol synthesis type-I polyketide synthase PpsC	92.31
PM0082941	PPE family protein	90.62	PM0082959	Phthiocerol synthesis polyketide synthase type-I PpsC	92.25
PM0082942	SDR family NAD(P)-dependent oxidoreductase	93.24	PM0082960	Polyketide synthase	92.31
PM0082943	PPE family protein	90.62	PM0082961	Phthiocerol synthesis polyketide synthase type-I PpsC	92.27
PM0082944	Polyketide synthase Pks12	92.67	PM0082962	Phthiocerol synthesis polyketide synthase type-I PpsC	92.31
PM0082945	Polyketide synthase	91.81	PM0082963	Phenolpthiocerol synthesis type-I polyketide synthase PpsC	92.31
PM0082946	PPE family protein PPE5	88.52	PM0082947	Phenolpthiocerol synthesis type-I polyketide synthase PpsC	92.34
PM0082948	Polyketide synthase PKS	93.24	PM0082967	Type-I polyketide synthase	91.84
PM0082949	Amino acid adenylation domain-containing protein	93.51	PM0082968	Polyketide beta-ketoacyl synthase	92.91
PM0082950	Uncharacterized protein	93.72	PM0082969	PPE family protein	100
PM0082951	Uncharacterized protein	93.95	PM0082970	Carrier domain-containing protein	93.50
PM0082952	Uncharacterized protein	93.33	PM0082971	Polyketide synthase	93.16
PM0082953	Phthiocerol/phenolphthiocerol polyketide synthase type-I PpsC	92.31	PM0083102	Fatty acid synthase	94.49
PM0082954	Phthiocerol synthesis polyketide synthase type-I PpsC	92.31	PM0083104	Probable fatty acid synthase FAS	94.11
PM0082964	Phenolphthiocerol synthesis type- PpsC	92.31	PM0083105	Type 1 polyketide synthase	91.86
PM0082965	Phenolpthiocerol synthesis type-I polyketide synthase PpsC	92.44	PM0083106	Fatty acid synthase	93.57
PM0082966	Polyketide synthase	92.52	PM0083107	Fatty acid synthase	91.92
PM0083113	PPE family protein	94.65	PM0083108	Probable fatty acid synthase FAS	91.92
PM0083114	Polyketide synthase PKS	89.36	PM0083133	Fatty acid synthase	92.99
PM0083115	PPE family protein	98.72	PM0083110	Type 1 polyketide synthase	97.89
PM0083116	PPE family protein	92.83	PM0083110	PPE family protein	91.76
PM0083117	Uncharacterized protein	98.31	PM0083112	PPE domain-containing protein	94.65
PM0083118	Uncharacterized protein	94.65	PM0083126	Non-ribosomal peptide synthetase	91.84
PM0083119	PPE family protein	95.02	PM0083130	Non-ribosomal peptide synthetase	98.10
PM0083120	Uncharacterized PPE family protein PPE54	94.65	PM0083131	amino acid adenylation domain-containing protein	94.49
PM0083121	PPE domain-containing protein	94.64	PM0083132	amino acid adenylation domain-containing protein	91.76
PM0083122	Peptide synthetase, putative	90.71	PM0083125	Non-ribosomal peptide synthetase	91.84
PM0083123	Peptide synthetase Nrp	91.85	PM0083134	PPE family protein	94.65
PM0083124	Probable protein synthetase nrp	94.65	PM0083127	Peptide synthetase	94.49
PM0083109	Putative FAS	96.88	PM0083128	Probable peptide synthetase Nrp (peptide synthase)	98.12
PM0083129	Putative peptide synthetase NRP (peptide synthase)	98.12			

PMDB: Protein Model Database, FAS: Fatty acid synthase

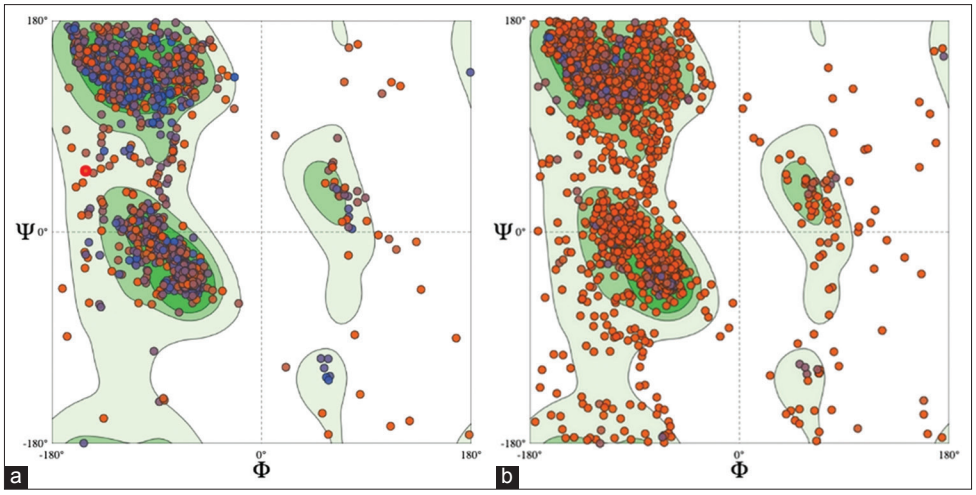

Fig. 1: Ramachandran plot analysis of modeled proteins; (a) Preferred model with minimum of >85% confidence score; (b) Rejected model with <85% confidence score

plot was used as the quantitative analysis of the reliability of the modeled proteins [22-24]. The result is a score of favorability called Ramachandran favored region. The Ramachandran favored also denoted in this study as confidence score for each model of each protein was stored. For the next procedure, the Ramachandran favored score was used for screening. Only those models with high confidence scores were taken into consideration [2].

Model submission

The 3D models thus predicted were submitted to the PMDB (http://srv00.recas.ba.infn.it/PMDB/main.php) which is a public resource database aimed at storing manually built 3D protein structures. The database is designed to provide access to models published in the scientific literature together with validating experimental data [1,2].

RESULTS

Building homology model

A total of 100 different protein sequences belonging to *M. tuberculosis* were retrieved from the UniProt database. All the sequences were subjected for BLASTn analysis in the NCBI tools, to identify significant match. Among the 100 sequences, only 69 protein sequences had 80% similarity match to the known protein structures in the PDB website. This suggests that significant portion of the proteins does not have known structures and is important to construct models of those proteins, for further applications. All the 100 sequences were analyzed to identify whether they are reported drug targets according to DrugBank database website. A total of 70 entries were identified as potential drug targets and thus playing important role in drug discovery and development. All the 100 sequences were subjected for homology model construction using the SWISS-Model online tool. The web tool has generated multiple models 1~5 different models for each entries. The best model for each protein was selected using Ramachandran plot analysis.

Ramachandran plot validation

Ramachandran plot analysis was employed as the quantitative analysis to predict the reliability of the modeled proteins, based on Ramachandran favored scores that are obtained by the MolProbity inbuilt within the SWISS-Model online tool. The predicted models were considered significant, only if the percentage of Ramachandran favored regions was above 85%. Hence, among the multiple models that were generated for each protein, the protein showed highest percentage of residues in the Ramachandran favored regions. Fig. 1 shows the Ramachandran plot analysis of a preferred model with >85% favored region and also a least preferred model with <85% Ramachandran favored region. A total of 69 protein models demonstrated significant score in Ramachandran plot and hence were selected for further processing.

Submission to PMDB

The 69 protein models that were selected from Ramachandran plot analysis were submitted to an online database, that is, PMDB (https://bioinformatics.cineca.it/PMDB/) for public availability to access for research purposes. The details of the constructed models and their PMDB entry ID are summarized in Table 1.

CONCLUSION

This study aimed at construction of computational 3D protein structures of *M. tuberculosis* using homology modeling. From the initial selected 100 proteins of the organism, a total of 69 proteins were successfully modeled with significant confidence score based on Ramachandran plot analysis. These modeled 3D structures were made available to the public through the PMDB database for computational protein models. Hence, in this study, the 3D structures of potential drug target proteins in *M. tuberculosis* were predicted and submitted for public access. This can be significantly useful for drug discovery and development of targeted drugs, specific to drug resistance, and strain specificity. The major problem in the treatment of tuberculosis is the rapid development of resistance and the varying sensitivity among different strains of the organism. Using the similar homology modeling approach, the drug resistance and sensitivity issues could be solved. This provides advantage to the structure-based computational drug design studies, on *M. tuberculosis* organism, aiding to developing an effective drug variant to overcome the current challenges faced by health care.

ACKNOWLEDGMENT

The authors thank the management of St. Joseph's College (Autonomous), Bengaluru, for supporting this research.

AUTHORS' CONTRIBUTIONS

All authors have significantly contributed toward completion of this work and construction of this manuscript.

FUNDING SOURCE

No funding source to support this research work.

CONFLICTS OF INTEREST

No known conflicts of interest.

REFERENCES

1. Sreenivas A, Imran S, Ravi L. Elucidation of computational 3D models of protein drug targets for *Colletotrichum falcatum* a fungal plant pathogen causing red rot of sugarcane. Biomed Pharmacol J 2020;13:627-33.
2. Jindam D, Ravi L, Krishnan K. Construction of computational protein data base by homology modeling for the aquatic pathogen *Perkinsus marinus* for targeted drug design and development. Res J Pharm Technol 2018;11:2203-8.
3. Feolo M, Helmberg W, Sherry ST, Maglott DR. NCBI genetic resources supporting immunogenetic research. Rev Immunogenet 2000;2:461-7.
4. Patel N, Prajapati N, Patel K, Patel R, Kalasariya H. Sequence Homology, Primer Designing and Homology Modeling Prediction by *In Silico* Pursuit. New Delhi: Microbiology in Services of Mankind; 2019.
5. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. The SWISS-MODEL repository and associated resources. Nucleic Acids Res 2009;37:387-92.
6. Kopp J. The SWISS-MODEL repository of annotated three-dimensional protein structure homology models. Nucleic Acids Res 2004;32:230D-4.
7. Fitzgerald D, Haas D. *Mycobacterium tuberculosis*. In: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. United Kingdom: Churchill Livingstone; 2005. p. 2852-86.
8. Talbot EA, Raffa BJ. *Mycobacterium tuberculosis*. In: Molecular Medical Microbiology. 2nd ed., Vol. 3. United States: Academic Press; 2014. p. 1637-53.
9. Ennassiri W, Jaouhari S, Sabouni R, Cherki W, Charof R, Filali-Maltouf A, *et al*. Analysis of isoniazid and rifampicin resistance in *Mycobacterium tuberculosis* isolates in Morocco using GenoType® MTBDRplus assay. J Glob Antimicrob Resist 2018;12:197-201.
10. Velayati AA, Farnia P, Hoffner S. Drug-resistant *Mycobacterium tuberculosis*: Epidemiology and role of morphological alterations. J Glob Antimicrob Resist 2018;12:192-6.
11. Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, *et al*. *Mycobacterium tuberculosis* complex genetic diversity: Mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol 2006;6:1-17.
12. Levine DM, Dutta NK, Eckels J, Scanga C, Stein C, Mehra S, *et al*. A *Tuberculosis* ontology for host systems biology. Tuberculosis 2015;95:570-4.
13. Sinha R, Rahul A. Breast *Tuberculosis*. Indian J Tuberc 2019;66:6-11.
14. Delogu G, Sali M, Fadda G. The biology of *Mycobacterium tuberculosis* infection. Mediterr J Hematol Infect Dis 2013;5:70.
15. Tsolaki AG, Gagneux S, Pym AS, de la Salmoniere YO, Kreiswirth BN, Van Soolingen D, *et al*. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. J Clin Microbiol 2005;43:3185-91.
16. Caulfield AJ, Wengenack NL. Diagnosis of active *Tuberculosis* disease: From microscopy to molecular techniques. J Clin Tuberc Other Mycobact Dis 2016;4:33-43.
17. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-science approach with life scientist friendly interfaces. BMC Bioinformatics 2010;11:548.
18. Canese K, Weis S. The bibliographic database. In: The NCBI Handbook. Bethesda, MD: National Center for Biotechnology Information; 2013.
19. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, *et al*. UniProt : The universal protein knowledgebase. Nucleic Acids Res 2004;32:D115-9.
20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403-10.
21. Waterhouse A, Bertoni M, Bienert S Studer G, Tauriello G, Gumienny R, *et al*. SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res 2018;46:W296-303.
22. Arg E. MolProbity Ramachandran analysis. Proteins 2003;437:180.
23. Hooft RW, Sander C, Vriend G. Objectively judging the quality of a protein structure from a Ramachandran plot. Bioinformatics 1997;13:425-30.
24. Hollingsworth SA, Karplus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. Biomol Concepts 2010;1:271-83.