CIBTech Journal of Zoology ISSN: 2319–3883
Online, International Journal, Available at http://www.cibtech.org/cjz.htm
2025 Vol.14, pp.43-49/Debangshi and Gowri
***Research Article** (Open Access)*

COMPUTATIONAL ANALYSIS TO PREDICT STRUCTURE AND FUNCTION OF TWO HYPOTHETICAL PROTEINS, ML0472 AND MLPM_0472 OF *MYCOBACTERIUM LEPRAE* COMPLEX

Debangshi Dasgupta, Gowri Devi K*
Swami Shraddhanand College, University of Delhi, Alipur, Delhi -110036 India
**Author for Correspondence: gowrizoology1@gmail.com*

ABSTRACT
The aim of this study was to focus on prediction of structure and function of two hypothetical proteins ML0472 and MLPM_0472 whose functions are uncharacterized by the use of computational methods and bioinformatics. A sequence of proteomic software's was used for predicting the possible structure along with a detailed structural analysis on the basis of disordered regions, GRAVY, aliphatic index, transmembrane helices, etc. DeepGOWeb webservice was used for predicting protein functions based on protein sequence using DeepGOPlus method. Results of this study showed ML0472 has 243 bp of gene length with 81 amino acids in sequence and MLPM_0472 has 90 amino acids. The computational analysis of protein ML0472 predicts its role in metabolism and of MLPM_0472 more in cellular processes than metabolism. Also, proteomic analysis has shown that it has a moderately positive instability index shows that it is a stable protein with a huge number of aliphatic chains. HARP database acts as a diversified tool to check the possible mutations in multi drug resistant patients. Hence, the DNA of *M. lepare* strain isolated from patients can be sequenced, and the possible mutations can be predicted by this HARP database. Our study has attempted to fill the gaps in the medical literature by proposing the structure and function of these two hypothetical proteins ML0472 and MLPM_0472 and the generated data may assist in the development of novel drug to treat leprosy.

***Keywords**: Mycobacterium leprae, hypothetical proteins, ML0472, MLPM_0472*

INTRODUCTION
Mycobacterium leprae is a gram positive, rod shaped acid fast bacterium, surviving exclusively within host cells which causes leprosy also known as Hansen's disease (Harald *et al.*, 2011). It is one of the most puzzling and complex microorganism as it forms a complex with *M. lepromatosis* called *M. leprae* complex which contains fewer protein-coding genes and large number of pseudogenes to cause the infection. Leprosy is a neglected tropical disease (NTD) which still occurs in more than 120 countries, with more than 200 000 new cases reported every year. During 2023, a total of 182 815 new cases were reported globally, corresponding to a new case detection rate of 22.7 per million population. The number of new cases detected globally was 5% higher than in 2022 (174 094) (Global leprosy (Hansen disease) update, 2023). Nowadays, the MDT based control measures for treating leprosy are mainly intended to prevent the drug resistant *M. leprae* from spreading (Maeda *et al.*, 2001). Despite the effectiveness of multidrug therapy (MDT), the slow reduction in new cases indicates ongoing transmission within communities (WHO Global Leprosy Update, 2018) and it remains a significant global health concern (Sakshi *et al*., 2021; Sengupta 2018; Rao Suneetha 2018).
The first genome sequence of a strain of *M. leprae*, was completed in 1998 (Cole *et al*., 1998). The genome sequence of *M. leprae* originally isolated in Tamil Nadu, India was completed in 2013. Its genome of 3,268,203 bp comprises approximately 1,604 proteins and contains 1,116 pseudogenes (Cole *et al*., 2001; Paul *et al*., 2015). They have annotated and classified all these genes into various functional categories. The complete genome contains about 360 conserved hypothetical and 141 unknown proteins

(Vissa and Brennan 2001). The hypothetical proteins are so named because it is unclear whether their genes encode actual proteins. These are also referred to as "uncharacterized" or "unknown" proteins (Sivashankari and Shanmughavel 2006). Sanmukh *et al*., 2011 studied the prediction of structure and function of hypothetical proteins in *M. Leprae* by the use of computational methods and bioinformatics. As of January 2022, the NCBI protein database contained 301 conserved hypothetical proteins of *Mycobacterium leprae* (Ojo *et al.,* 2022). Therefore, it is clear that the different genomes remain unrevealed at time and the proper structural and functional annotation of hypothetical proteins may lead to better understanding of the complex protein cascades, protein-protein interaction or networks in several biological systems (Loewenstein *et al.,* 2009). This understanding will be helpful in both diagnostic and therapeutic purposes (Lubec *et al*., 2005). Protein characterization is important for determining the current state of any protein, which has significant implications to several biological processes. But there are major shortcomings of experimental methods used to characterize the proteins of various organisms like time consuming, cost effective (Alterovitz *et al*., 2005). The Computational or in silico approaches provide a viable solution to these problems.

Different bioinformatics algorithms were used to conduct structural and functional annotations of a hypothetical protein ML-1369 in *M. leprae*, in which the primary sequence analysis was conducted using Protparam server and the secondary structure of the query protein was analyzed using SOPMA and PSIPRED server; the 3D model for the protein was generated using SWISS-MODEL server by homology modelling method (Paul *et al*., 2015). Hardik *et al*., 2020 studied an *in-silico* approach for function prediction of hypothetical proteins of *M. leprae*TN by applying ten diverse tools like BLASTP, InterProScan, COGniter, protFUN, CDD, TMHMM, Phyre meta server, PFP.

In this study, for the first time the structure and function of two hypothetical proteins, ML0472 and MLPM_0472 of *M. leprae* complex were analysed using *in silico* methods. Different computational tools were used to generate the three-dimensional (3D) model of the protein. Moreover, the functional analyses were also conducted to propose the possible functions of the protein and research in drug design planning.

MATERIALS AND METHODS

Searching for the Protein sequence of ML0472 and MLPM_0472

NCBI is the directory of all sequences discovered till date and can be used to find possible gene or protein sequences. Mycobrowser is also dedicated software for Mycobacterial sp. through which the gene and protein sequence of *M. leprae* and *M. lepromatosis* can be determined.

Structural analysis of the genome

A sequence of proteomic software's was used for predicting the possible structure along with a detailed structural analysis on the basis of disordered regions, GRAVY, aliphatic index, transmembrane helices, etc.

Software	*Function*
RaptorX	3D structure prediction by deep learning
Phyre2	Alignment and structural predictions
I-TASSER	Hierarchy approach for structural and structure based functional approach
Rosetta	3D Structure predictor of biological macromolecules
Alpha fold	Predict 3D structure of protein from its amino acid sequence
Swiss Model	Database of annotated 3D structure models by homology modelling
PONDR	Predicts natural disorder region
TMHMM	Predicts transmembrane helices in protein

Functional analysis of the genome

DeepGOWeb is a web service used for predicting protein functions based on protein sequence using DeepGOPlus method. It uses deep convolutional neural networks to learn sequence features and combines predictions with sequence similarity based predictions. After submission of FASTA it predicted possible function in cellular and biological processes with molecular function. VICM pred is a web server that aid in broad functional classification of proteins of bacteria into virulence factors, information molecule, cellular process and metabolism. The VICM pred server uses support vector machine (SVM) based algorithm having patterns, amino acid and dipeptide composition of bacterial protein sequences and overall accuracy of this server is 70.75%.

Drug Susceptibility testing for M. leprae

HARP is a database that predicts multi drug resistance mutations in leprosy. This server has targeted on the following drugs.

Drug	Drug target	Mutations
Dapsone	Dihydropteroate Synthase (DHPS)	folp1 gene
Rifampicin	RNA Polymerase	rpoB, rpoC, etc
Ofloxacin	DNA gyrase	GyrA

RESULTS AND DISCUUSION

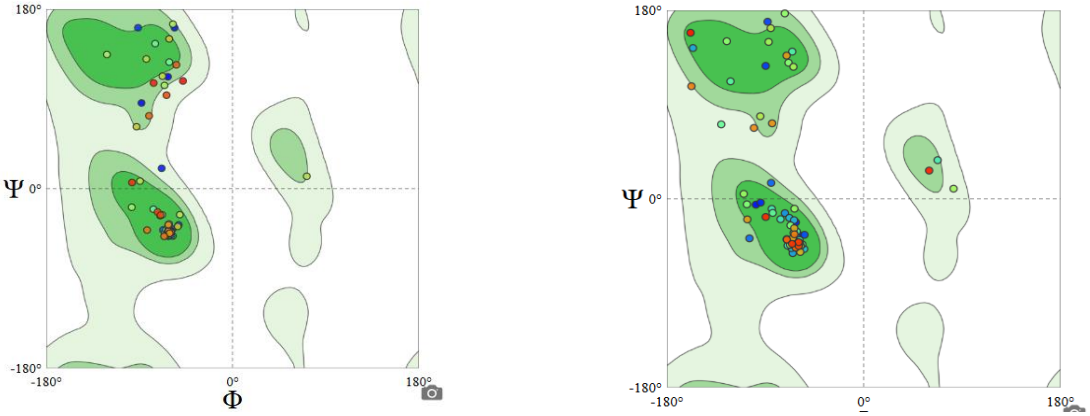
After getting protein sequence of ML0472 and MLPM_0472 from Mycobrowser, i.e.,
>Mycobacterium leprae TN|ML0472|ML0472
VFKLLGFGWLTDVGLFGYLAAWEVFVSILGEVVAATNPDELLEALATITDDQWYREFTASTVIAL
LAYFNSPCSACRPLL and
>Mycobacterium lepromatosis Mx122A|MLPM_0472|MLPM_0472
MFKRLGFSWRTDVALPSSLAAWEVFVSTLGQMVVATNSNEPFEELSTMTDDRRNREFTAPAVIAL
LAYFMFVLQCMSTIGVMRRGIHSW, we have used this sequence for proteomic analysis. After using various softwares we found our proteins 3D conformation and its various integrity using Alpha fold and Swiss Model (Figure 1). Ramachandran plot with most of its amino acids lying in right handed alpha helix and beta sheet region, while only single amino acid in left handed alpha helix region (Figure 2). The disordered segment of ML0472 was found between region [33] – [42] and of MLPM_0472 between [34]-[48] using PONDR (Figure 3). ML0472 showed a moderate to low number of transmembrane helices with some above 0.6 P value in [20] region of segment and MLPM_0472 have high transmembrane helices with 1 P value using TMHMM (Figure 4). ML0472 predicted to have role in metabolism of *M. leprae* and MLPM_0472 have role in cellular processes of *M. lepromatosis* using VICM pred (Figure 5), which can be further examined for drug resistance after sequencing through wet lab experiments and using the HARP database.

Elimination of this neglected tropical disease i.e., leprosy is important to avoid spread of microbial resistance. Sanmukh *et al*., 2011 studied the prediction of structure and function of hypothetical proteins in *M. Leprae* by the use of web tools (CDD-BLAST, INTERPROSCAN, PFAM and COGs). The 3-D structures of the 160 functionally resolved hypothetical proteins were constructed using PS2 server (Protein structure prediction server). Paul *et al*., 2015 designed a study to predict the 3D structure and biological function of the hypothetical protein ML-1369 of *M. Leprae* and retrieved the complete protein sequence of this protein in FASTA format successfully from the Uniprot database, this protein was predicted to be consisting of 231 amino acids, with a molecular weight of 25329.1 Daltons. Proteomics has provided considerable insight into the hypothetical proteins of *M. leprae* complex. The structural analysis has predicted about its various integrities along with its role in cellular and metabolic functions.

HARP database acts as a diversified tool to check the possible mutations in multi drug resistant patients.

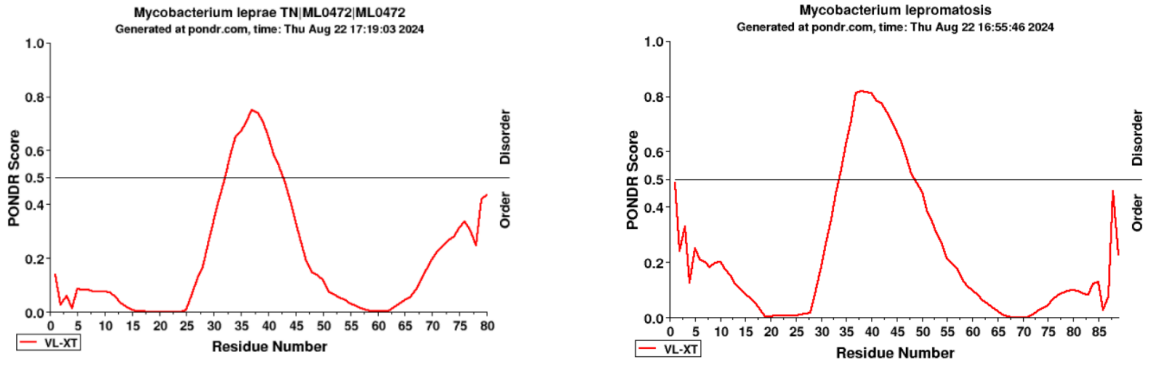
(A) **(B)**

Figure 1. 3D Structures of ML0472 (A) and MLMP0472 using RaptorX



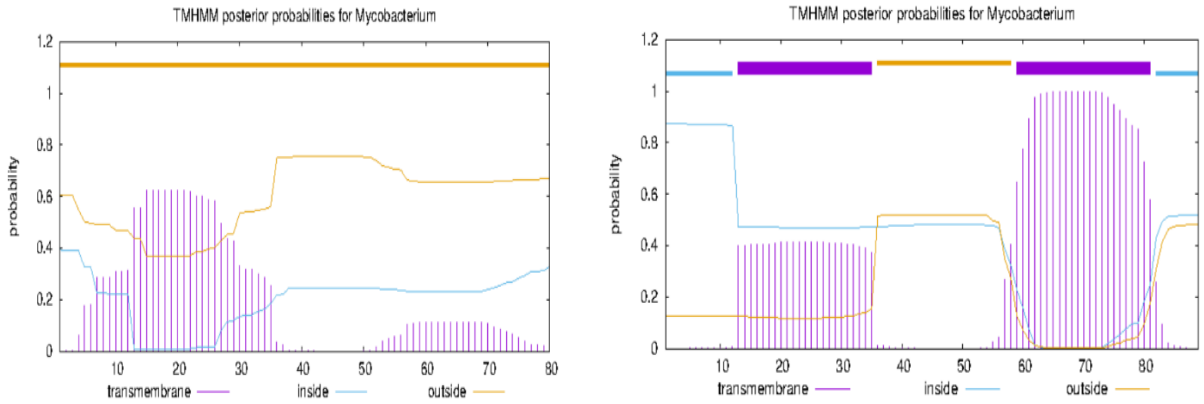
(A) **(B)**

Figure 2. Ramachandran plots for detailed analysis of the amino acids using Swiss Model: ML0472 (A) and MLMP0472.



(A) **(B)**

Figure 3. The disordered segments of ML0472 (A) and MLMP0472 predicted using PONDR



(A) **(B)**

Figure 4. Analysis of transmembrane helices using TMHMM: ML0472 (A) and MLMP0472

Score of Different Functional Class	
Function	Score
cellular Process	-0.88212519
Information Molecule	-4.1113104
Metabolism	1.7513775
Virulence factors	-2.7598615

Score of Different Functional Class	
Function	Score
cellular Process	-0.58607483
Information Molecule	-5.6864437
Metabolism	-0.86302094
Virulence factors	-1.6085242

(A) **(B)**

Figure 5. Functional characterization of ML0472 (A) and MLMP0472 using VICM Pred

Hence, the DNA of *M. leprae* strain isolated from patients can be sequenced, and the possible mutations can be predicted by this HARP database. In the near future, proteomics is anticipated to evolve into a vital instrument in diagnosing leprosy and pinpointing therapeutic markers (Harald *et al*., 2011). Expression and functional activation of pseudogenes still remains as a question for the researchers (Sakshi *et al*., 2021).

CONCLUSION

The dissemination of *Mycobacterium leprae*, the causative agent of leprosy, continues to be a lasting concern globally. Although various interventions have been implemented over the years, outcomes have been inconsistent, highlighting the critical necessity for discovering novel biomarkers for this ailment. A multitude of investigations have examined the application of proteins as possible diagnostic and predictive markers by employing proteomic analysis. This methodology provides significant understanding of the interaction of cellular processes, supplementing genomics and conventional biochemical methods. The generated data may assist in the development of novel drug to treat leprosy.

The present work stands out in the fact that it is the first report of structure and functions of hypothetical proteins, ML0472 and MLPM_0472 of *M. leprae* complex using computational methods and adds to the scientific knowledge on ML0472 and MLPM_0472 hypothetical proteins.

ACKNOWLEDGEMENT

The authors are grateful to the Principal, Swami Shraddhanand College, University of Delhi, for providing laboratory facilities for the completion of the present research study.

Conflict of interest: The authors declare that there is no conflict of interest.

REFERENCES

Alterovitz G, Afkhami E, Ramoni M (2005). Robotics, automation, and statistical learning for proteomics. *In:* New Developments in Robotics Research (Liu JX, ed). New York: *Nova Science Publishers,* Inc 2005; pp. 217-252. *https://doi.org/10.1002/pmic.200600060*.

Cole ST, Brosch R, Parkhill J (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature,* **393**(6685) 37-544. *10.1038/31159*.

Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D (2001). Massive gene decay in the leprosy bacillus. *Nature,* **409**(6823) 1007-*1011*. *10.1038/35059006.*

Global leprosy (Hansen disease) update (2023). Elimination of leprosy disease is possible -Time to act!

Harald G, Wikera, Gisele G. Tomazellaa. and Gustavo A. de Souzaa (2011). A quantitative view on *Mycobacterium leprae* antigens by proteomics. *Journal of Proteomics*, **74** 1711-1719. *https://doi.org/10.1016/j.jprot.2011.01.004*.

Hardik S, Aishwarya J, John JG (2020). An *in silico* approach for function prediction of hypothetical proteins of *Mycobacterium laprae* TN. *Proceedings of the National Conference on Innovations in Biological Sciences (NCIBS), January 10, 2020* 322-9.

Loewenstein Y, Raimondo D, Redfern OC, Watson J, rishman D, Linial M, Orengo C, Thornton J, Tramontano A (2009). Protein function annotation by homology-based inference. *Genome Biology,* **10**(2) 207. *https://doi.org/10.1186/gb-2009-10-2-207*.

Lubec G, Afjehi-Sadat L, Yang JW, John JP (2005). Searching for hypothetical proteins: theory and practice based upon original data and literature. *Progressive Neurobiology,* **77**(1) 90-127. *https://doi.org/10.1016/j.pneurobio.2005.10.001*.

Maeda S, Matsuoka M, Nakata N, Kai M, Maeda Y, Hashimoto K, Kashiwabara Y (2001). Multidrug resistant *Mycobacterium leprae* from patients with leprosy. *Antimicrobial Agents Chemotherapy,* **45**(12) 3635-3639.

Ojo O, Williams DL, Adams LB, Lahiri R (2022). *Mycobacterium leprae* transcriptome during In Vivo growth and Ex Vivo stationary phases. *Frontiers in Cellular and Infection Microbiology,* **11**, 817221. *https://doi.org/10.3389/fcimb.2021.817221*.

Paul S, Saha1 M, Talukdar SN, Rajbongshi S, Akhand RN, Md. Aminul Islam (2015). Computational Approaches Predict the Reliable Three Dimensional (3D) Structure and Possible Involvement of a Hypothetical Protein of *Mycobacterium leprae*, ML-1369 in Its Cell Division. *Journal of pure and applied Microbiology,* **9**(2) 919-926.

Rao PN, Suneetha S (2018). Current Situation of Leprosy in India and its Future Implications. *Indian Dermatology,* **9**:83-89. *10.4103/idoj.IDOJ_282_17*.

Sakshi G, Devesh S, Anjana G, Shripad AP, Deepa B (2021). Insights into *Mycobacterium leprae* Proteomics and Biomarkers-An Overview. *Proteomes,* **9**(1) 7. *10.3390/proteomes9010007*.

Sanmukh SG, Paunikar WN, Ghosh TK (2011). Computational approach for structure and functionality search for hypothetical proteins in *Mycobacterium laprae*. **3**(5) 281-296.

Sengupta U (2018). Elimination of leprosy in India: An analysis. *Indian Journal of Dermatology Venereology and Leprology,* **84** 131-136. *10.4103/ijdvl.IJDVL_1070_16.*

Sivashankari S, Shanmughavel P (2006). Functional annotation of hypothetical proteins–A review. *Bioinformation*, **1**(8) 335-338. *https://doi.org/10.6026/97320630001335*.

Vissa VD, Brennan PJ (2001). The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set. *Genome Biol*, **2**(8):1023.1-1023.8. *https://doi.org/10.1186/gb-2001-2-8-reviews1023*.

World Health Organization (2019). Global Leprosy Update, 2018: Moving towards a Leprosy Free World; Weekly Epidemiological Record; *World Health Organization: Geneva, Switzerland*, **94** 389-412.

CIBTech Journal of Zoology ISSN: 2319–3883
Online, International Journal, Available at http://www.cibtech.org/cjz.htm
2025 Vol.14, pp.43-49/Debangshi and Gowri
***Research Article** (Open Access)*