



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 14, Issue 3, March 2025**

# Identification of Arresting Capsular A-Glucan Biosynthesis in Mycobacterium Tuberculosis by using Computer-Aided Methods

Lavanya Rumandla<sup>1</sup>

Department of Chemistry, Vivekananda Government Degree College (Autonomous), Vidyanagar, Telangana, India<sup>1</sup>

**ABSTRACT:** The devastating illness, Tuberculosis (Tb) is a major global health concern, affected by the Mycobacterium tuberculosis (Mtb). The Capsular glucan Synthase (glgA) belongs to the glycosyl transferase group 2 family and is targeted to identify active ligand molecules to inhibit the Capsular  $\alpha$ -glucan biosynthesis. Using Modeller 9.13, and the homology simulation model, the 3D model was generated. The resulting protein structure was verified, and the active site was located utilizing prerequisites. The protein 3D structure consists of 11 sheets and 15 helices. At the active site, the grid was constructed, and the definder database was used for virtual screening. Most of the ligand molecules consist of azole moieties and their derivatives as pharmacophore groups in their structures. The amino acids from virtual screening are VAL37 crucial for binding to the ligand molecules. The lead compounds with advantageous ADME characteristics can restrict the formation of Capsular glucan in M. tuberculosis. The azole structural scaffolds are found in ligand molecules of the BL2, BL4, and BL6, docking complexes.

**KEYWORDS:** Mycobacterium tuberculosis, Capsular  $\alpha$ -glucan biosynthesis, Modeller, virtual screening, ADME.

**ABBREVIATIONS:** Tb-Tuberculosis, Mtb-Mycobacterium Tuberculosis, 3D-Three dimensional, ADME-Absorption, distribution, metabolism, and excretion, BL-Binding ligand

## I. INTRODUCTION

The etiological origin of Tuberculosis, Mycobacterium tuberculosis (Mtb), was discovered by Robert Koch ("[The etiology of tuberculosis by Dr Robert Koch]," 1982) more than 140 years ago and still continues to threaten public health around the world. The global number of 1.3 million deaths officially announced by WHO impacted by the COVID-19 pandemic in 2020 was almost doubled the number (0.68 million) infected by TB related to HIV/AIDS (WHO Tuberculosis Report 2021, n.d.). In the present study, the target protein glgA, functionally a Capsular glucan synthase protein involved in the formation of D-glucan in the capsule formation (Sambou et al., 2008). The inactivation of glgA has inhibited the formation of capsular  $\alpha$ -D-glucan, whereas disruption of glgC was reducing the formation of glucan and glycogen levels by half (Cywes et al., 1997), indicating that the glgA was inhibited is essential in capsular  $\alpha$ -D-glucan biosynthesis. As a result, confirming that the glgA protein is a viable therapeutic target for designing new chemical entities that impede capsule bio-synthesis for the survival of Mtb in the host cells.

## II. METHODS AND MATERIALS

### Homology modelling

Understanding the target protein structure is important for comprehending its biology. Due to the non-availability of "X-ray crystal and NMR structure of target protein" (Offman et al., 2008), homology modelling is applied to generate a 3D structure. The target protein, glgA Mycobacterium tuberculosis with a sequence of 389 amino acid chain length is obtained from Universal Protein Knowledge Base (UniProtKB) ("UniProt: The Universal Protein Knowledgebase in 2021," 2021) in FASTA format. Templates are selected using template search tools, Phyre2 (Kelley & Sternberg, 2009) (protein homology/analogy recognition engine) for structural similarity and Jpred4 (Cole et al., 2008), (Drozdetskiy et al., 2015) for secondary structure.



**Sequence Alignment and model building**

A better 3D model will result from strong sequence alignment having target and template proteins. The resulting alignment file caters for the information of the residues identical, similar residues, and different residues using the ClustalW (Cambridge, UK) server tool. The sequences of the target and the template are aligned pair-wise. The glgA protein comparative model was created through the automated homology modelling tool MODELLER(Sali et al., 1995),(Janson et al., 2019). The twenty models were created by employing MODELLER 9.13.

**Optimization and validation of the 3D model**

Energy optimization is a crucial step in stabilizing the created 3D model. For structural optimization(Jorgensen & Tirado-Rives, 1988), the protein preparation module Maestro 9.0.111 (Schrodinger LLC, New York) is utilized. The model quality of glgA protein was enhanced by locPREFMD(Feig, 2016) (local protein structure refinement via molecular dynamics), a web server tool. The refined 3D model is examined using the PROCHECK and ProSA servers as validation tools. The Ramachandran plot in PROCHECK(Laskowski et al., 1996) explains the stereochemical quality, by analysing the angles at which amino acid residues are phi and psi. PROSA(Wiederstein & Sippl, 2007), (Protein Structural analysis) server validated the protein model for local compatibility in terms of energy.

**Prediction of Active site**

Finding the active region is key for ligand-based drug design. The CASTp (Computed Atlas of Surface Topology of Protein(Jorgensen & Tirado-Rives, 2005),(Dundas et al., 2006) and Schrodinger suite's site map module software are used to predict the active site of the glgA protein. The CASTp server application uses the Protein's molecular surface and the solvent-accessible surface model to detect the binding pockets. The sitemap estimates the total site score with excellent precision. The Sitemap lists out the protein binding regions, their hydrophobicity, hydrogen bond donors, and the interactions among the template and target proteins.

**Virtual Screening**

The systematic three steps in the Virtual Screening method are 1) Grid generation, 2) Ligand preparation and 3) Docking. Virtual screening is carryout on the grid region to find the lead compounds for the target protein.

Generation of Grid: On the energy-minimized three-dimensional protein structure, the grid is created around the area of the active site. A grid having specifications (30x30x30) Å is created using the Schrodinger Suit of Grid Gen module. The whole active site grooves, including all binding opportunities for different ligand conformations are covered by the Grid box.

Preparation of lead molecules: On selecting the ligand molecules from many chemical databases, including the definder database in structure data file (SDF) format, and procured using LigPrep module (Schrodinger Suite 2010) with OPLS 2005(Friesner et al., 2006) force field. The ligand preparation technique produces 4 conformers with lower energy per each ligand at a predetermined  $P^H$  of  $7.0 \pm 2.0$ . Virtual screening is applied to obtain conformers with the lowest energy.

Docking: In the virtual screening process, at the receptor active site location, ligand structures from LigPrep are employed. The Glide (Grid-based Ligand Docking with Energetics) (Ballante et al., 2021) a Virtual Screening workflow from Schrodinger software, which has the following scoring protocols with a 10% output each: HTVS-high - throughput virtual screening, SP-standard precision, and XP-extra precision. For further studies, the docked ligands are considered with the highest Glide energy and Glide Score(Friesner et al., 2006).

**III. RESULTS AND DISCUSSION****Structural assessment and validation**

The glgA, glycogen transferase group 1 family protein is involved in Capsular glucan Biosynthesis. Capsule formation is a crucial stage in the existence of Mycobacterium TB, allowing the pathogen to adapt itself in an unfavourable environment for a prolonged period in the host.

The glgA protein and 3V template were then aligned pair by pair by ClustalW. 88% structural similar residues were shown, a perfect alignment of template and protein pair by pair. MODELLER 9.13 was used to create 40 homogeneous models initially. For further research, the model with lowest the Modeller parameter was chosen. The produced protein model was made stable by energy minimization in Schrodinger Suite - Protein preparation wizard. The model's fidelity

was greatly increased by utilizing locPREFMD, a web server tool, for Molecular Dynamics. The results show that the stability characteristics of the glgA structure significantly improved after applying molecular dynamics. Energy is refined and structural changes were done to improve stereochemical quality.

The Ramachandran plot of phi against psi values for the model amino acid residues of the protein, which was used to validate the stereo-chemical quality shown Out of 389 amino acids, the results indicate that 300 residues in the most preferred zone, 26 in the extra permitted region, and 8 residues in the generously allowed section justified the high stereo-chemical properties of the protein model. ProSa server evaluated the Z-score acquired from X-ray crystallography and NMR-spectroscopy of the Protein Data bank, displays the aquired Z-score of -7.79, a negative value that indicates the glgA model has a high quality structure. Generated glgA protein 3D model is more reliable, according to amino acid residues below the base line with negative energies.

### Protein secondary structure and active site prediction

The protein secondary structure, which has 11  $\alpha$ -helices and 12  $\beta$ -sheets as predicted by the PDBsum server (Table 1, and Table 2). CASTp server suggested the putative active site, whereas NCBI-BLAST server predicted the conserved domain. The 14 helix-helix interactions present in glgA protein observed. The non-bonding interactions are a key feature for protein structural stabilization observed in biological macromolecules. The glgA protein three-dimensional structure was stabilized using molecular dynamics simulation. The CASTp server locates and delineates concave surface areas on the three-dimensional design of glgA protein and binding pockets 1,2, and 3 are indicated by red, orange, and blue colours respectively. PyMOL software was utilized to visualize the optimized 3D model. The amino acids from ARG8 to LEU60 were identified at the protein's active site in the further screening experiments.

**Table2.** Helices representation in glgA protein secondary structure

Helix no	Initial residue	End residue	Residue number	length	Residue sequence
1	Thr22	Leu28	7	13.07	TELVAYL
2	Ala69	Leu71	3	-	ADL
3	Trp89	Tyr102	14	20.18	WYTALAGHLAAILY
4	Glu114	Leu116	3	-	EPL
5	Gly126	Ala141	16	23.51	GYQVSTWVEQTAVLAA
6	Ala147	Val155	9	13.90	AMRNDMLRV
7	Arg186	Leu194	9	13.73	RTGSVLAEL
8	Arg210	Lys212	3	-	RQK
9	Val215	Ala222	8	12.61	VVHLVTAA
10	Val244	Leu251	8	12.23	VRVAVAE
11	Ile268	Ala279	10	15.67	IGQLREILSA

**Table2.** Analysis of Secondary structure of Protein glgA obtained from PDBsum server tool. This table consists the data of 11 alpha helices with Amino acid chain length, residue number and residue sequence of the protein.

**Table3.** Sheets representation in secondary structure of glgA protein

Number	Starting residue	Ending residue	Number residues	Residue Sequence
1	Val4	Leu7	4	VAML
2	Val33	His36	4	VDVH
3	Val82	His86	5	VVHSH
4	His102	Thr104	4	HVLT
5	Val145	Val148	4	VIAV
6	Val166	Val168	3	VHV
7	Met201	Val205	5	MAVFFV
8	Glan221	Cys235	5	QLVLC

9	Gly257	Ile261	5	GVFWI
10	Val280	Cys283	4	VFVC
11	Ala303	Ser307	5	AVVAS
12	Gly324	Val328	4	GSLV

**Table3.** The glgA protein secondary structure was obtained from the PDBsum server. The details shown in the table were 12  $\beta$ -sheets with their lengths and sequences.

#### Virtual screening

A grid was created with the dimensions of 32 X32 X 32 Å at the active site area of protein. The LigPrep module of the Schrodinger suite was employed to prepare the ligand structures for Virtual screening. The Schrödinger suite's ligand preparation module received 11987 compounds from the definder database, and 20497 molecules were obtained as output. This module applies hierarchically the filters, HTVS (high throughput virtual screening), then SP (Standard Precision), and XP (Extra Precision) in each stage having a default functional output of 10%. The top 6 complexes are arranged in order of their glide energy and glide score as virtual screening out of the 20 docked complexes.

It is observed that in table 4, BL1 to BL6 top ten ligand molecules were categorized according to the non-bonding interactions. The ligand molecules that were docked showed a consistent pattern of binding with heterocyclic ring structures containing nitrogen that was actively interacting with the amino acid VAL37 of the target protein. Most of the ligand molecules consist of azole moieties and their derivatives as pharmacophore groups in their structures. Prior research has demonstrated the importance of the heterocyclic azole moieties in the process of developing new TB drugs. These azole structural scaffolds are found in ligand molecules of the BL2, BL4, and BL6 docking complexes.

#### ADME Properties

The assessment of the ADME study is a pivotal point for subsequent pre-clinical research of drug development to boost drug discovery's success rate. In Table 5, the In-silico prediction of ADME attributes has been implemented for the docked complexes from BL1 to BL6. These pharmacokinetic parameters are calculated by comparing the pharmacokinetic features of pre-existing drugs through the QikProp module (Schrodinger LLC, 2010, New York). The results of this investigation imply that identified ligands have drug-like results due to their adherence to the allowed range.. In this contrast, the newly selected ligand molecules are being considered as a prospective therapeutic approach to TB by arresting the Capsular glucan biosynthesis. By comparing the residues' surface accessibility to that of free glgA protein, according to the findings, the amino acids VAL37 of glgA protein are involved in selective ligand binding and play a key role in inhibiting capsular glucan biosynthesis.

**Table5.** The ADME features of the ligands that have been prioritized

S. No	CNS	Mwt	DonorH B	Accepted HB	Q Plog Po/w	QPlog BB	%Human oral absorption	Rule of three	Rule of five
BL1	-2	390.458	4.0	8.5	1.636	-1.518	80.990	0	0
BL2	-2	310.371	1.5	9.0	0.918	-0.638	82.630	0	0
BL3	-1	257.296	2.25	3.25	1.020	-0.866	73.057	0	0
BL4	0	256.229	2.0	4.7	1.526	-0.494	89.734	0	0
BL5	-2	457.508	3.0	6.0	4.817	-1.749	100.000	1	0
BL6	-2	359.462	3.0	3.5	2.473	-1.109	87.694	0	0

**Table5.** The ADME attributes predicted by the Schrödinger suite's Quick Prop module.

#### IV. CONCLUSIONS

The 3D structure of the 389 amino acid glgA protein was analysed, and the structural features of the protein were derived using energy minimization. Virtual screening experiments were conducted in order to find and report the drug-like most effective lead inhibitor compounds of glgA protein to arrest the capsule glucan biosynthesis. The amino acids

VAL37 have a significant impact on protein-ligand binding. The lead compounds with acceptable ADME characteristics and a satisfactory docking score were found and reported. According to the findings, the azoles structural scaffolds were identified as drug-like lead potential molecules in the development of anti-tuberculosis drugs.

**ACKNOWLEDGEMENTS:** The authors RL thankful to the HEAD, Department of Chemistry, and the Principal, Vivekananda Government degree college for providing facilities to carry out paper work.

#### REFERENCES

1. The etiology of tuberculosis by Dr Robert Koch]. (1982). Salud publica de Mexico, 24(3), 241–257. <http://europepmc.org/abstract/MED/6815810>
2. Ali, S. A., Hassan, M. I., Islam, A., & Ahmad, F. (2014). A review of methods available to estimate solvent-accessible surface areas of soluble proteins in the folded and unfolded states. Current Protein & Peptide Science, 15(5), 456–476. <https://doi.org/10.2174/1389203715666140327114232>
3. Cole, C., Barber, J. D., & Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Research, 36(Web Server issue), 197–201. <https://doi.org/10.1093/nar/gkn238>
4. Cywes, C., Hoppe, H. C., Daffé, M., & Ehlers, M. R. (1997). Nonopsonic binding of Mycobacterium tuberculosis to complement receptor type 3 is mediated by capsular Chttps://doi.org/10.1128/iai.65.10.4258-4266.1997
5. Drozdetskiy, A., Cole, C., Procter, J., & Barton, G. (2015). JPred4: A protein secondary structure prediction server. Nucleic Acids Research, 43. <https://doi.org/10.1093/nar/gkv332>
6. Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., & Liang, J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Research, 34(Web Server issue), W116–8. <https://doi.org/10.1093/nar/gkl282>
7. Feig, M. (2016). Local Protein Structure Refinement via Molecular Dynamics Simulations with locPREFMD. Journal of Chemical Information and Modeling, 56(7), 1304–1312. <https://doi.org/10.1021/acs.jcim.6b00222>
8. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., & Mainz, D. T. (2006). Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. Journal of Medicinal Chemistry, 49(21), 6177–6196. <https://doi.org/10.1021/jm051256o>
9. Kelley, L. A., & Sternberg, M. J. E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nature Protocols, 4(3), 363–371. <https://doi.org/10.1038/nprot.2009.2>
10. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of Biomolecular NMR, 8(4), 477–486. <https://doi.org/10.1007/BF00228148>
11. Offman, M. N., Tournier, A. L., & Bates, P. A. (2008). Alternating evolutionary pressure in a genetic algorithm facilitates protein model selection. BMC Structural Biology, 8(1), 34. <https://doi.org/10.1186/1472-6807-8-34>
12. Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., & Karplus, M. (1995). Evaluation of comparative protein modeling by MODELLER. Proteins, 23(3), 318–326. <https://doi.org/10.1002/prot.340230306>
13. Savojardo, C., Manfredi, M., Martelli, P. L., & Casadio, R. (2020). Solvent Accessibility of Residues Undergoing Pathogenic Variations in Humans: From Protein Structures to Protein Sequences. Frontiers in Molecular Biosciences, 7, 626363. <https://doi.org/10.3389/fmolb.2020.626363>
14. UniProt: the universal protein knowledgebase in 2021. (2021). Nucleic Acids Research, 49(D1), D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
15. WHO Tuberculosis report 2021. (n.d.).
16. Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research, 35(suppl\_2), W407–W410. <https://doi.org/10.1093/nar/gkm290>





INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details