IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

IN-SILICO VIRTUAL SCREENING STRATEGIES TO FIND MAZF6 ENDORIBONUCLEASE PROTEIN INHIBITORS FOR MYCOBACTERIUM TUBERCULOSIS

[1]Dr. Rumandla Lavanya*, [2]Dr Pagudala Adavi Raju*

[1] Dr Rumandla Lavanya: Associate Professor, Department of Chemistry, Vivekananda GDC(A),Vidyanagar, Hyderabad, Telangana.

[2] Dr Pagudala Adavi Raju: Associate Professor, GDC, Malkajgiri, Hyderabad, Telangana**.**

1

**E-mail: rumandla572010@gmail.com*

Abstract: Tuberculosis is a worldwide threat people with low immune system and poverty, which is caused by Mycobacterium tuberculosis bacterium. Mazf family proteins causes virulence due to the adaptable nature of the pathogen in the host body. Homology modelling techniques are employed in drug development such as template section methods such as BLAST, PHYRE2, and JPred4 server tools. Sequence alignment done by CLUSTALW, and model building using MODELLER 10.4. model validation done by Ramachandran plot and ProSA techniques. Protein preparation carried out through Schrodinger suite protein preparation wizard, and active site prediction by CASTp, ligand preparation, grid generation followed by virtual screening and pharmacological studies ADME properties are assessed for best target molecules. The lead molecules are identified from D1 to D4 with common pharmacophore, benzyl amine functional group.

Key words: Template, MODELLER, Active site, ADME properties, Lead molecules.

INTRODUCTION

Tuberculosis is a major threat to public wellness caused by the bacteria Mycobacterium tuberculosis. During the COVID-19 epidemic, “the number of people misdiagnosed and untreated has escalated, resulting in Tb death and community transmission of infection, with some time-lag, increasing a greater number of people infecting TB,[1]” according to the WHO 2022 report[2]. To combat the spreading tuberculosis etiology and emerging multi-drug-resistant strains[3][4] new targets for inhibiting Mycobacterium tuberculosis are required with new mechanisms of action. In this present study Mazf family protein[5] adaptable nature towards survival of the drastic existing conditions of the Mtb bacteria. MazF6 protein has been shown crucial for virulence and physiology. The MazF6 protein work in concert to let Mtb adapt to situations like oxidative stress, food deprivation, and drug exposure[6]. The current study emphasizes how crucial MazF6 endoribonucleases inhibited with lead molecules and identify the prospective targets for the treatment of tuberculosis.

METHODOLOGY

2.1. Homology modeling

Computational techniques[7] are employed in drug development to overcome the limitations of traditional procedures[8]. The UniProt Knowledgebase (UniProtKB)[9],[10] server delivered the data in FASTA format. The sequences were analysed using BLAST[11],[12], PHYRE2[13], and the PSI-BLAST module of the JPred4[14] server tools. The template was chosen based on the following criteria: sequence similarity, query coverage, low E-value, protein fold identity, and similar secondary structure.

2.2. Alignment And Model Generation

We identified conserved identical, similar, and non-identical areas by using CLUSTALW to align numerous sequences between the amino acid sequences of the biological target and the reference template[15]. MODELLER 10.4[16] is used to create three-dimensional models of the MazF6 protein. The model was developed by constrained optimization of the protein sequence alignment[17]. Thirty models were initially constructed, and the one with the lowest MOF (Modeller objective function) was chosen for further refinement testing.

2.3. Structural Optimization of MazF6 and Validation

The developed 3D model stability can be accomplished by 3D model optimization. The produced 3D model's energy conservation is performed using imperf[13] and Protein preparation wizard, Schrödinger suite by, OPLS (Optimal Potential for Liquid Simulations) force field 2005. PROCHEK scores as the primary assessment criterion, the stereochemical quality of the data used to generate the model is improved, bringing the data closer to what is expected for experimental structures. To validate the improved MazF6 protein model, the PROCHECK and ProSA server programs were used. The protein model's stereochemical quality, local model quality, and overall model quality was all assessed and validated. PROCHEK[18] generates Ramachandran plots to assess amino acid stereochemistry based on phi(Φ) and psi(ψ) angles. The Protein structural Analysis (ProSA) server[19] program examines the energy and local compatibility of the protein structure in comparison to others of similar amino acid length that of PDB.

2.4. Prediction of MazF6 protein active site

Potential active sites are found where the ligand interacts with the protein. The MazF6 protein's active site is identified as the binding pockets on the surface of the protein. “The identification and size characterization of surface pockets and occluded are the first stages in designing ligands based on protein structure”. The active location is often the biggest pocket or cavity”[20],[18].The CASTp server[21]. The CASTp detects the protein structure’s binding cavities and pockets[22]. The validation process for binding site recognition and assessment can assist researchers in locating binding sites with high confidence and predicting their draggability. SiteMap also aids in lead optimization by offering insight into receptor-ligand interactions and recommending viable ways for modifying lead molecules to improve receptor complementarity[23].

2.5. Docking and Virtual screening investigations

The Glide module of the Schrödinger software is used to dock and screen ligands from several small molecule databases against the biological target as part of the structure-based drug discovery technique[24] A grid was established in the MazF6 protein's binding domain area. Virtual Screening[24] uses ligands from structural databases. The LigPrep module of the Schrodinger suite[22] of Virtual Screening is used to optimize the stereochemistry, ring confirmers, and ionization states of the ligands. The docking process for the Glide employs three repetitions of the Virtual Screening protocols[25], namely the HTVS, SP mode, and XP mode. The docked ligands are ranked using both the Glide score and the Glide energy.

2.6. Quikprop properties

The Schrödinger suite's Qikprop module is used to look into the ADME characteristics of certain ligand molecules. It's crucial to understand in-silico Qikprop (Absorption, Distribution, Metabolism, and Excretion) characteristics to determine how similar lead compounds are to drugs[26],[27].

RESULTS ANALYSIS

3.1. Data analysis of target sequence and building MazF6 model

The target protein, MazF6, is a endoribonuclease protein and a chain length of 234 amino acid residues. It belongs to MazF family. The MazF6 protein is functionally involved in the virulence of the pathogen even in the drastic conditions.The target protein’s FASTA sequence was obtained through the Universal Protein

Knowledge Base (UniprotKB) server tool. The template was discovered utilizing the MazF6 protein's main sequence in FASTA format as an input file for several servers such as BLAST, Jpred 4, and PHYRE2. The template, PDB ID 1NE8A has been found with 39.95% sequencing identity, 97% query coverage, and low e-values, as shown in table 1. The target and homologous template proteins sequence similarity was determined using the ClustalW server programme. The MazF6 protein has a better pairwise similarity with the 1NE8A template protein as 29.7% identical, and 39.9% similar. The template coordinates and alignment files were submitted to MODELLER 10.4 and 30 models were created, the model with the least MOF (2463.48560) being selected for further investigation.

Table 1. Selection of Template results by the comparison of different search tools

The data base tool	Parameters	E-score	PDB code of template
Blast	Similarity in sequence	6e-04	1NE8A
Jpred4	Prediction of secondary structure	5e-17	1NE8A
PHYRE2	Protein fold recognition	3e-20	1NE8A

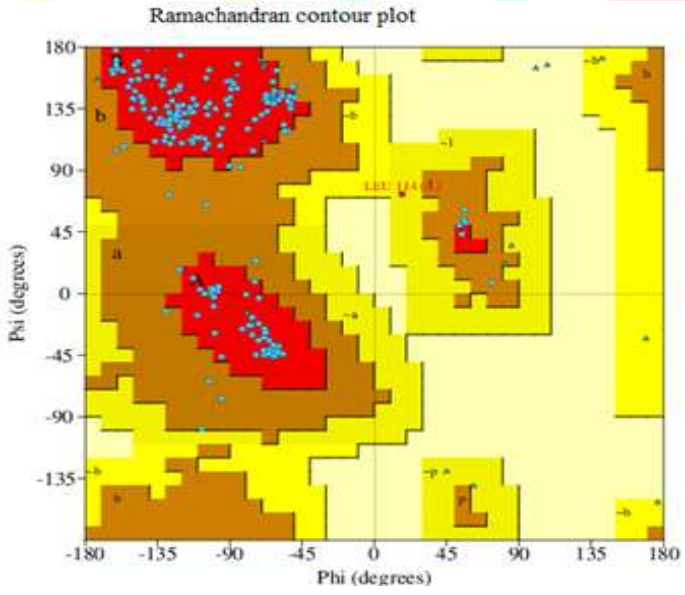
Table 1. Template selection done by comparing the various search tools such as Blast, Jpred 4, and PHYRE2 with lowest e-value.

3.2. Energy minimization and model validation

Schrodinger Suite Protein preparation wizard is utilized to decrease the protein energy. Protein structure local stereochemistry, which is where the backbone is conserved but molprobidity scores are considerably improved.Making use of the Ramachandran Plot, which is shown in Figure 1(a) as a contour plot of the backbone dihedral angles [phi (Φ) versus psi (ψ)]. 88.2% (314) of residues are located in the most preferred region, followed by 9.8% (35) in the allowed region, 1.4% (5) in the generously allowed zone, and 0.6% (2) in the restricted region, per the plot data (Figure 1(b)). As a consequence, the 3D model created is of great quality. The ProSA energy map explains why this model's local quality is acceptable and trustworthy. The ProSA energy map explains the protein's simulated structure local quality. The black spot in Figure 2(a) represents the model protein's Z-score is -3.61. The obtained Z-score of the protein structure was measured by x-ray or NMR crystallography. The bulk of the amino acid residues are below the baseline in the ProSA local quality graph (Figure 2(b)). Validation studies conducted on the examined 3D structure show the model to be of high quality.

Figure 1. Ramachandran contour diagram of MazF6 protein

a)



b)

Regions in the RC plot	Number of residues	percentage
Most favoured region	**176**	**92.1%**
Additionally allowed region	**14**	**7.3%**
Generously allowed region	**1**	**0.5%**
Disallowed regions	**0**	**0.0%**
Non-glycine residues and non-proline residues	**191**	**100%**
End residues (excluding Gly and pro)	**2**	
Number of glycine residues	**20**	
Number of proline residues	**21**	
Total number of residues	**234**	

Figure 1(a) shows the Ramachandran contour diagram of MazF6 protein. **(b)** shows the statistics of Ramachandran plot.

Figure 2. The ProSA analysis of MazF6 protein

a) b)

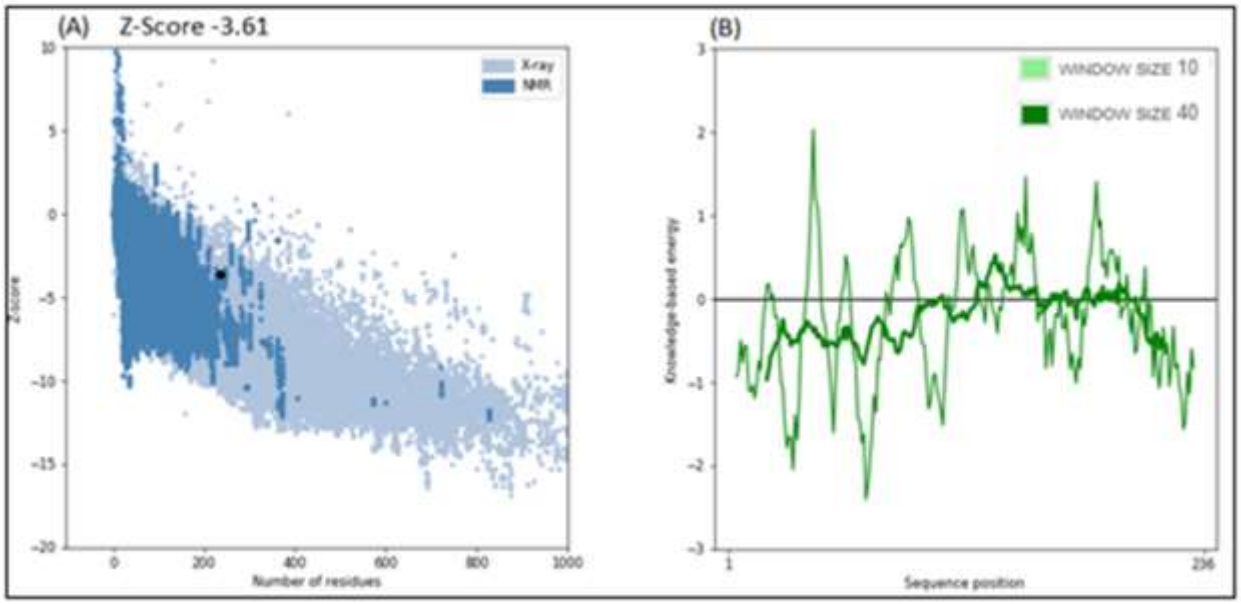


Figure 2 a) ProSa graph of MazF6 protein. Z-score of MazF6 model is -3.61, which is indicative of good overall quality pf protein. **b)** energy profile graph shows the amino acids of MazF6 are falling below the baseline, a good quality protein.

3.3. Secondary structure prediction and active site prediction for MazF6 protein

The validated model of MazF6 protein 3D structure was visualized using PYMOL software, and comprised up of 8 helices, 10 sheets, and one hairpin. The PDB-Sum server programme was utilised to access the secondary structure. The structure strengthened by 10 helix-helix interactions. Knowing the binding location provides essential information in drug development. The binding site protocols CASTp, and SiteMap used to detect the active site of MazF6 protein. The binding site area of conserved area of MazF group family is 102 - 148 amino acid residues. CASTp, the active site prediction server, evaluated the binding cavities depended on the hydrophobicity, resulting two cavities, shown in Figure 3 as cavity 1 and cavity 2 regions. The SiteMap recognizes the hydrogen bond acceptors, donors, and hydrophilic regions on the protein surface. In Association with all of the above data from several binding site prediction algorithms such as CASTp, SiteMap, and BLAST conserved domain analysis. A grid was constructed at the binding site region of the MazF6 protein spanning 99 to 150 residues.

Figure 3. The putative binding pockets of the MazF6 protein

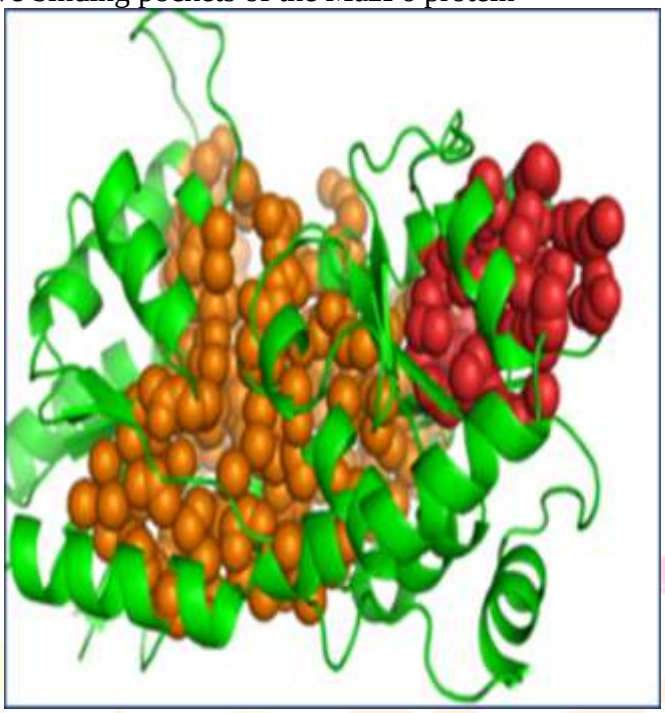
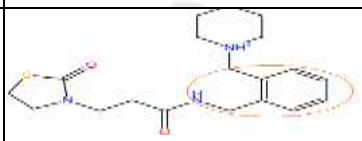
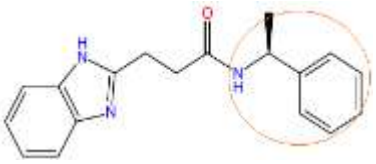


Figure 3. The putative binding cavities of MazF6 protein of CASTp, visualized through PyMoL. The binding Pockets 1 is in yellow color and 2 was shown red in color.

3.4. Virtual Screening

A grid was created with 36X36X36 Å dimensions using the amino acids centroid of the residues of the protein's binding region. The Virtual Screening was performed on the protein grid and ligand library of chemical substances. The ligand preparation module was used to prepare 11000 ligand molecules from the Hitdefinder database. For each ligand, four stereo-confirmers with distinct chiralities were retained. As reliable confirmers 0f 20430 and these created confirmers were run through the HTVS, SP and XP modes with a 10% filtering ratio in each stage of docking. A total of 20 binding complexes were obtained, with glide scores ranging from -13.53 to -8.09. Table 4.4 shows a selection of 4 docked ligand molecules categorized by XP Glide score and Glide energy from D1 to D4 are selected for the study. Biovia Discovery Studio has been employed to visualize the binding interactions (Table 2). In the present study, ligand molecules with benzyl amine groups as pharmacophores.

Table 2. Binding interaction of docking complexes

Ligand No	Structure of the molecule	Glide energy(K.cal /mol)	XP Glide score	Intermolecular interaction	Distance (Å)
D1	 3-Thiazolidinepropanamide, 2oxo-N-(2-(1-piperidinyl methyl)	-40.11	-13.53	ASN126:H– O17:D1 SER132:OG–H16:D2	1.80 1.83
D2	 1H-Benzimidazole-2-	-40.49	-12.39	ASN126:H - N1:D2 SER132:OG–H8:D2 Pi-Cation HIS143:ND1-D2	1.87 2.05 6.84

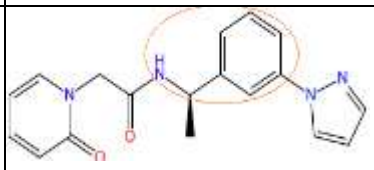
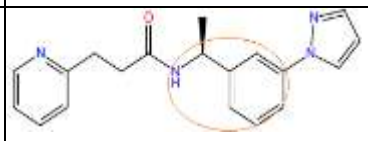
	propanamide, N-(1-phenylethyl)				
D3	 **1(2H)-pyridine acetamide, 2-oxo-N-(3-(1H-pyrazole-1-yl)**	**-45.15**	**-8.09**	**ASN126:H-O25:D3 SER132:OG –H7:D3**	**1.80 1.96**
D4	 **2-pyridinepropanamide, N-(3-(1H-pyrazole-1-yl) phenyl ethyl)**	**-39.19**	**-9.79**	**ASN126:H–N16: D4 SER132: OG-H5: D4**	**2.00 1.89**

Table 2. The binding interactions of the MazF6 Docking complexes. The D1 to D4 ligands molecules are represented based on their glide scores, and energies.

3.5. Quikprop properties

The antagonistic reactivity of an inhibitor to a protein receptor does not necessarily predict its therapeutic efficacy. As a result, ADME studies are crucial for drug development since they help determine whether or not it is appropriate to provide inhibitors to a biological system. Furthermore, the majority of failed clinical trial drugs are frequently caused by inhibitors that have undesirable ADME properties and substantial unfavorable impacts on the biological system. The QuikProp module is used to examine the ADME properties of specific ligand molecules. compounds are within the permitted range of ADME parameters considered as anti-tuberculosis lead compounds (Table 3). The CNS values that fell within the range of -2 to 0. Results of the Blood/Brain barrier patrician coefficient values are from -2.44 to -0.36, the Donor hydrogen bond is from 2.0 to 4.5, The acceptor hydrogen bond is from 3 to 4.5, and molecular weights of ligands are from 561.165 to 652.152 are within the permissible range. The Lipinski rule of three and rule five values are within the desirable range of ADME properties. The selected lead molecules were shown to have permissible pharmacokinetic features and were also potent inhibitors of the MazF6 protein.

Table 3. ADME properties of selected ligands from the docking studies.

S. No	Glide score	CNS	Mwt	Donor Hydrogen bond	Accepted hydrogen bond	Qplog Po/w	Qplog BB	Oral absorption % in human	Rule of three	Rule of Five
D1	**-13.53**	**-2**	**569.155**	**2.0**	**3.0**	**-2.36**	**-2.44**	**41.365**	**0**	**0**
D2	**-12.39**	**-1**	**652.152**	**4.0**	**4.0**	**5.65**	**-0.70**	**67.62**	**1**	**0**
D3	**-8.09**	**0**	**561.165**	**3.0**	**3.0**	**-0.56**	**-0.36**	**72.61**	**0**	**0**
D4	**-9.79**	**-1**	**572.201**	**4.0**	**4.5**	**-3.06**	**-1.71**	**40.095**	**1**	**0**

Table 3. The ADME attributes are predicted by the Schrödinger suite's Quick Prop module. The following allowed ADME value ranges are: Oral absorption in humans (%): > 80 high, 25 is least; The Rule of Three (3) is followed by the Rule of Five (4). Mol wt. (130 to 725); CNS: -2(inactive) +2(active); HB of the donor: (0.0 to 6.0); Qplog po/w: (-2 to 6.5); Qplog BB: (-3 to 1.2); Accept HB (2.0 to 20). % of Human oral absorption:>75%.

HB – Hydrogen bond, Po/w – Partition coefficient of Octanol of Water, BB – Brain blood barrier

Conclusion

The MazF6 endoribonuclease family protein involved in the virulence of the Mtb. A 3D model is created and validated using the necessary comparative modelling rules. The grid is built at the MazF6 protein's estimated active site area for docking experiments with the Hit definder database.4 Ligand molecules are selected From D1 to D4 based on their Glide score. The benzyl amine group is a common pharmacophore

present in the D1 to D4 docked complexes and have common binding with the ASN126, and SER132, amino acid residues in the MazF6 protein selected on the binding site area (Table 2). The Benzyl amine pharmacophore with the permissible ADME pharmacokinetic properties is significantly proven as drug target molecules. Hence the lead molecules in this study can efficiently inhibit the virulence and simultaneously can control the *Mtb*.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the Head of the Department of Chemistry and the Principal of the University College of Science at Osmania University for facilitating the research work.

References:

[1] T. Dlangalala, A. Musekiwa, and T. Mashamba-Thompson, "Impact of COVID-19 on TB diagnostic services at primary healthcare clinics in eThekwini district, South Africa," *Sci. Rep.*, vol. 13, no. 1, p. 16645, 2023, doi: 10.1038/s41598-023-43589-7.

[2] "Global Tuberculosis reporT 2022," 2022. [Online]. Available: http://apps.who.int/bookorders.

[3] A. Koul, E. Arnoult, N. Lounis, J. Guillemont, and K. Andries, "The challenge of new drug discovery for tuberculosis," *Nature*, vol. 469, no. 7331, pp. 483–490, 2011, doi: 10.1038/nature09657.

[4] H. M. Berman *et al.*, "The Protein Data Bank.," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, Jan. 2000, doi: 10.1093/nar/28.1.235.

[5] P. Carroll, A. C. Brown, A. R. Hartridge, and T. Parish, "Expression of Mycobacterium tuberculosis Rv1991c using an arabinose-inducible promoter demonstrates its role as a toxin.," *FEMS Microbiol. Lett.*, vol. 274, no. 1, pp. 73–82, Sep. 2007, doi: 10.1111/j.1574-6968.2007.00842.x.

[6] A. Gupta, "Killing activity and rescue function of genome-wide toxin–antitoxin loci of Mycobacterium tuberculosis," *FEMS Microbiol. Lett.*, vol. 290, no. 1, pp. 45–53, Jan. 2009, doi: 10.1111/j.1574-6968.2008.01400.x.

[7] A. V. Sadybekov and V. Katritch, "Computational approaches streamlining drug discovery," *Nature*, vol. 616, no. 7958, pp. 673–685, 2023, doi: 10.1038/s41586-023-05905-z.

[8] A. B. Gurung, M. A. Ali, J. Lee, M. A. Farah, and K. M. Al-Anazi, "An Updated Review of Computer-Aided Drug Design and Its Application to COVID-19," *Biomed Res. Int.*, vol. 2021, 2021, doi: 10.1155/2021/8853056.

[9] "UniProt: the universal protein knowledgebase in 2021," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D480–D489, Jan. 2021, doi: 10.1093/nar/gkaa1100.

[10] "UniProt: the Universal Protein Knowledgebase in 2023," *Nucleic Acids Res.*, vol. 51, no. D1, pp. D523–D531, Jan. 2023, doi: 10.1093/nar/gkac1052.

[11] C. Shen *et al.*, "A generalized protein-ligand scoring framework with balanced scoring, docking, ranking and screening powers," *Chem. Sci.*, vol. 14, no. 30, pp. 8129–8146, 2023, doi: 10.1039/d3sc02044d.

[12] S. F. Altschul *et al.*, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.," *Nucleic Acids Res.*, vol. 25, no. 17, pp. 3389–3402, Sep. 1997, doi: 10.1093/nar/25.17.3389.

[13] R. Vadija *et al.*, "Identification of small molecular inhibitors for efflux protein Rv2688c of Mycobacterium tuberculosis," *Biotechnol. Appl. Biochem.*, vol. 65, no. 4, p. 608—621, Jul. 2018, doi: 10.1002/bab.1647.

[14] A. Drozdetskiy, C. Cole, J. Procter, and G. Barton, "JPred4: A protein secondary structure prediction server," *Nucleic Acids Res.*, vol. 43, Apr. 2015, doi: 10.1093/nar/gkv332.

[15] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "CLUSTAL W: improving the sensitivity of

progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.," *Nucleic Acids Res.*, vol. 22, no. 22, pp. 4673–4680, Nov. 1994, doi: 10.1093/nar/22.22.4673.

[16] B. Webb and A. Sali, "Comparative Protein Structure Modeling Using MODELLER.," *Curr. Protoc. Bioinforma.*, vol. 47, pp. 5.6.1-32, Sep. 2014, doi: 10.1002/0471250953.bi0506s47.

[17] G. Janson, A. Grottesi, M. Pietrosanto, G. Ausiello, G. Guarguaglini, and A. Paiardini, "Revisiting the 'satisfaction of spatial restraints' approach of MODELLER for protein homology modeling," *PLOS Comput. Biol.*, vol. 15, no. 12, p. e1007219, Dec. 2019, [Online]. Available: https://doi.org/10.1371/journal.pcbi.1007219

[18] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, "PROCHECK: a program to check the stereochemical quality of protein structures," *J. Appl. Crystallogr.*, vol. 26, no. 2, pp. 283–291, Apr. 1993, doi: https://doi.org/10.1107/S0021889892009944.

[19] M. Wiederstein and M. J. Sippl, "ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins," *Nucleic Acids Res.*, vol. 35, no. suppl_2, pp. W407–W410, Jul. 2007, doi: 10.1093/nar/gkm290.

[20] R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein, and J. M. Thornton, "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.," *J. Biomol. NMR*, vol. 8, no. 4, pp. 477–486, Dec. 1996, doi: 10.1007/BF00228148.

[21] W. Tian, C. Chen, X. Lei, J. Zhao, and J. Liang, "CASTp 3.0: computed atlas of surface topography of proteins," *Nucleic Acids Res.*, vol. 46, no. W1, pp. W363–W367, Jul. 2018, doi: 10.1093/nar/gky473.

[22] J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, and J. Liang, "CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues.," *Nucleic Acids Res.*, vol. 34, no. Web Server issue, pp. W116-8, Jul. 2006, doi: 10.1093/nar/gkl282.

[23] T. Halgren, "New method for fast and accurate binding-site identification and analysis.," *Chem. Biol. Drug Des.*, vol. 69, no. 2, pp. 146–148, Feb. 2007, doi: 10.1111/j.1747-0285.2007.00483.x.

[24] W. L. Jorgensen and J. Tirado-Rives, "Potential energy functions for atomic-level simulations of water and organic and biomolecular systems.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 19, pp. 6665–6670, May 2005, doi: 10.1073/pnas.0408037102.

[25] W.-H. Shin, L. Heo, J. Lee, J. Ko, C. Seok, and J. Lee, "LigDockCSA: Protein–ligand docking using conformational space annealing," *J. Comput. Chem.*, vol. 32, 2011, [Online]. Available: https://api.semanticscholar.org/CorpusID:13798429

[26] Z. Ahmad, S. Sharma, and G. K. Khuller, "In vitro and ex vivo antimycobacterial potential of azole drugs against Mycobacterium tuberculosis H37Rv.," *FEMS Microbiol. Lett.*, vol. 251, no. 1, pp. 19–22, Oct. 2005, doi: 10.1016/j.femsle.2005.07.022.

[27] A. J. Lucas, J. L. Sproston, P. Barton, and R. J. Riley, "Estimating human ADME properties, pharmacokinetic parameters and likely clinical dose in drug discovery," *Expert Opin. Drug Discov.*, vol. 14, no. 12, pp. 1313–1327, 2019, doi: 10.1080/17460441.2019.1660642.