Research Article

BIOMEDICAL Journal of Scientific & Technical Research

ISSN: 2574 -1241

DOI: 10.26717/BJSTR.2025.61.009581

Structure Prediction, Characterization, and Functional Annotation of Functional Protein of *Mycobacterium Basiliense*: An *In Silico* Approach

Mubeen Ali Niaz[1]*, Muhammad Zohaib Niaz[2], Hafiz Shahid Hussain[3] and Muhammad Adnan Shah Bukhari[1]

[1]*Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Pakistan*

[2]*Department of Bioscience and Technology, Emerson University, Pakistan*

[3]*Xinjiang Institute of Ecology & Geography, Chinese Academy of Sciences, China*

***Corresponding author:** Mubeen Ali Niaz, Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Multan, Pakistan

ARTICLE INFO

Received: March 21, 2025

Published: April 09, 2025

Citation: Mubeen Ali Niaz, Muhammad Zohaib Niaz, Hafiz Shahid Hussain and Muhammad Adnan Shah Bukhari. Structure Prediction, Characterization, and Functional Annotation of Functional Protein of *Mycobacterium Basiliense*: An *In Silico* Approach. Biomed J Sci & Tech Res 61(2)-2025. BJSTR. MS.ID.009581.

ABSTRACT

Mycobacterium basiliense is an emerging pathogenic species capable of thriving in human macrophages. To better understand its pathogenic mechanisms, this study presents an in-depth in silico analysis of the hemerythrin-like protein VDM89547.1 encoded in the *M. basiliense* genome. Physicochemical characterization revealed a 20.9 kDa, thermolabile protein with a highly helical secondary structure. Tertiary modeling and refinement generated an accurate 3D model (MolProbity score 1.13). Molecular docking with glutathione showed favorable binding, mediated through predicted hydrogen bonds and salt bridges across an extensive interface. A 1000 ps molecular dynamics simulation confirmed structural stability. Overall, these analyses provide the first insights into VDM89547.1's structure-function relationships. The thermolability and numerous putative binding sites suggest involvement in critical heat-sensitive molecular interactions within M. basiliense. Elucidating the protein's binding partners and clarifying its physiological role may reveal new antibiotic targets against this emerging human pathogen. This study lays the foundation for future experimental investigations of this novel protein.

Keywords: *Mycobacterium Basiliense; In Silico* and Hemerythrin Like Protein; Tuberculosis; Structure Prediction

Abbreviations: MSAS: Multiple Sequence Alignments; PAE: Predicted Alignment Error; MD: Molecular Dynamics; GRAVY: Grand Average of Hydropathicity; MB: Megabases

Introduction

Mycobacterium basiliense, a recent addition to the slow-growing mycobacteria family, is a species that belongs to the Mycobacterium genus [1]. This genus is mainly associated with pulmonary diseases, with Mycobacterium tuberculosis being the notorious culprit behind granulomatous pulmonary infections [2]. In contrast, *M. basiliense* stands out as a novel species that was initially isolated from respiratory samples of five patients, four of whom had existing pulmonary conditions. To comprehensively characterize this new species, various biochemical and molecular techniques were employed, including whole genome sequencing. While its biochemical traits align with those of *M. marinum* and *M. Ulcerans* [3], a significant distinguishing feature of *M. basiliense* is its ability to grow at 37°C, which is typically unfavorable for many mycobacterial species. Furthermore, *M. basiliense* has shown an ability to thrive within human macrophages but not in amoebae, indicating a propensity towards pathogenicity rather than an environmental lifestyle.

The complete genome sequencing of *M. basiliense* revealed a genome size of approximately 5.6 megabases (Mb) characterized by structural integrity and synteny similar to that of *M. marinum* while exhibiting minimal insertion sequence presence in its genome structure. Noteworthy are the numerous virulence factors found in the genome such as 283 PE/PPE surface-associated proteins constituting around 10% of the coding capacity alongside 22 non-ribosomal

peptide synthase clusters [1]. Interestingly enough, comparative genomic analysis conducted on six clinical isolates from the five affected patients revealed minor genetic variations differing only by up to two single nucleotide polymorphisms suggesting a probable common source of infection which further emphasizes the distinctiveness of *M. basiliense* as a taxonomic entity [4]. The focus of this study is to present a comprehensive analysis of the Hemerythrin-like protein VDM89547.1 found in *M. basiliense*. It encompasses predictions regarding its tertiary structure, distinctive characteristics, and functional attributes associated with ligand binding. This valuable information contributes to our comprehension of the protein's role and potential functions within the bacterium's biological framework. Furthermore, it could potentially yield insights with far-reaching implications, including drug development and enhancing our understanding of the bacterium's interactions with its surroundings.

Materials & Methods

Sequence Retrieval

The Hemerythrin-like protein, identified as VDM89547.1 in the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/), lacks a 3D structure in the Protein Data Bank (PDB) (https://www.rcsb.org/) [5]. Consequently, we focused on modeling the secondary and tertiary structures of 186 amino acids long protein from *M. basiliense*. Our goal was to characterize and provide functional annotation for this protein.

Physiochemical Properties

For the evaluation of the physicochemical attributes of the protein, we utilized the Expasy web-based computational tools. Specifically, we employed the ProtParam tool (https://web.expasy.org/protparam/) [6], which enabled us to predict various critical characteristics. These included the protein's potential for instability, aliphatic index, analysis of the amino acid composition, estimation of the theoretical isoelectric point (pI), and calculation of the Grand Average of Hydropathicity (GRAVY) score.

Secondary Structure Prediction

In the pursuit of identifying secondary structural elements, this study employs two reliable prediction methods. The first approach utilized is the self-optimized prediction with alignment (SOPMA) (https://npsa-prabi.ibcp.fr/NPSA/npsa_sopma.html), renowned for its robustness and ability to provide accurate estimations [7]. Additionally, the PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/#) program, known for its proficiency in secondary structure prediction, is also incorporated [8]. By combining these methods, a comprehensive and precise assessment of the secondary structure of the VDM89547.1 protein is ensured. This meticulous approach greatly enhances our understanding of the intricate three-dimensional structure of the protein, which plays a vital role in comprehending its functional properties and potential binding sites for ligands.

Tertiary Structure Modelling

We employed AlphaFold 2 to predict the tertiary structure of the hemerythrin-like protein from *M. basiliense*. The prediction was carried out using the Google Colab implementation of AlphaFold. Retrieved sequence was then input into the AlphaFold Colab notebook, which utilizes pre-trained AlphaFold 2 models. The notebook automatically generated multiple sequence alignments (MSAs) using the MMseqs2 tool, searching against the UniRef90 and MGnify databases. Template structures were identified from the PDB70 database. The AlphaFold model then processed this information through its neural network architecture, including the Evoformer blocks and structure module. The output included the predicted 3D coordinates of the protein structure, per-residue confidence scores (pLDDT), and a predicted alignment error (PAE) matrix. We visualized the resulting structure using ChimeraX and analyzed the confidence scores to identify regions of high and low prediction accuracy. This approach allowed us to generate a high-quality structural prediction for this previously uncharacterized protein from *M. basiliense*, providing insights into its potential function and evolutionary relationships.

Tertiary Structure Refinement and Validation

After carefully selecting the tertiary structure, we utilized the GalaxyWeb Refine Server (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) to further refine and optimize its three-dimensional arrangement [9]. This step was essential in improving the accuracy and reliability of the protein. To evaluate the quality of the modeled tertiary structure, we conducted a thorough analysis called the Ramachandran plot using PROCHECK (https://saves.mbi.ucla.edu/). This analysis offers valuable insights into the stereochemical quality of the protein structure by examining the distribution of dihedral angles of its amino acid residues [10]. For final validation, we turned to MolProbity analysis (http://molprobity.biochem.duke.edu/index.php), a platform that provides comprehensive validation and refinement capabilities [11]. Through this platform, we ensured that our protein model is accurate and reliable. To assess the overall quality of our tertiary model, we relied on Z-scores generated from ProSAweb (https://prosa.services.came.sbg.ac.at/prosa.php) [12]. These scores give us a quantitative measure of how well our protein structure fits within the expected value range. They assist us in evaluating its structural integrity and stability effectively.

Molecular Docking

To investigate the interaction between *M. basiliense* protein and Glutathione (GSH), molecular docking was employed. Docking is a computational technique used to predict the binding mode and affinity of a ligand (GSH, in this case) to a receptor (VDM89547.1). For this study, the ClusPro web server (https://cluspro.bu.edu/home.php), a widely utilized and reliable docking platform, was employed [13]. Before initiating the docking process, PDB files were prepared. Any water molecules, co-crystallized ligands, or non-standard residues

were removed by using the DockPrep tool ChimeraX software [14]. The structure was then energetically minimized to correct any structural irregularities or steric clashes that may affect the accuracy of the docking results. ClusPro employs a multi-step process that integrates global energy minimization, pairwise interaction energy estimation, and clustering algorithms to generate potential binding poses of the ligand within the binding pocket of the receptor protein.

The server provides a user-friendly interface, allowing for easy submission and retrieval of results. The receptor protein file in PDB format and the ligand structure in PDB were uploaded to the ClusPro server. Parameters such as grid dimensions, clustering threshold, and number of structures to retain were set based on default values and standard recommendations from the ClusPro documentation. Once the docking calculations were completed, the results were downloaded from the ClusPro server. The generated poses were ranked based on various scoring functions, including electrostatics, van der Waals interactions, and desolvation energy. The most energetically favorable binding poses were further analyzed for their interactions, orientation, and potential hydrogen bonding patterns between the protein and GSH.

Molecular Dynamic Simulation

Molecular dynamics (MD) simulations were performed with WebGRO, an online protein folding and stability analysis tool (https://simlab.uams.edu/) [15]. The refined tertiary structure model was submitted to the WebGRO server for aqueous solvation and dynamics simulation using default parameters. The protein was oriented and placed in the center of a cubic water box with edges at least 15 Å from the protein surface. The AMBER14SB force field was employed and the system charge was neutralized with NaCl ions at a physiologic 150 mM concentration. Energy minimization was carried out before the production MD phase. A 1000 ps simulation at 300 K temperature was executed, saving coordinates every 1 ps, resulting in 10,000 model snapshots. Key measurements were computed over the trajectory frames to evaluate stability. These included radii of gyration, root mean square deviation of backbone atoms, root mean square fluctuations of residues, solvent accessible surface area, and protein-water hydrogen bond occupancy. Trajectory visualizations were also generated. Collectively, these metrics quantified retention of the overall tertiary structure and local dynamic movements during the simulation timeframe [16].

Results and Discussion

Physiochemical Characterization

The amino acid sequence of VDM89547.1, found in *Mycobacterium basiliense*, was extracted in FASTA format and utilized as a reference sequence to calculate various physicochemical parameters (Table 1). The instability index of VDM89547.1, which registers at 42.60 (greater than 40), signifies its inherent instability as a protein. Furthermore, this protein exhibits an acidic nature, with a calculated isoelectric point (pI) of 5.43, and possesses a molecular weight of 20,947.80 Da. Notably, the VDM89547.1 protein boasts a higher aliphatic index, measuring at 90.27, indicating a pivotal factor contributing to its enhanced thermostability over a broad temperature spectrum. This protein is characterized as hydrophilic, suggesting a favorable propensity for interactions with water, a trait underscored by its lower grand average of hydropathicity (GRAVY) index value (-0.420), as detailed in the accompanying Table 1. Additionally, the amino acid composition, outlined in Table 2 and acquired through the Expasy ProtParam Tool, provides essential insights into the protein's constitution (Figure 1). This composition becomes instrumental in discerning active amino acid pockets, a valuable resource for potential drug-targeting endeavors aimed at this protein.

Table 1: Physiochemical parameters.

Physiochemical parameter	Value
Number of Amino Acid	186
Molecular Weight	20947.80
Theoretical isoelectric point (pI)	5.43
Aliphatic Index	90.27
Instability Index	42.60
Total No. of Negatively charged residues (Asp + Lys)	34
Total No. of Positively charged residues (Arg + Lys)	25
Grand Average of Hydropathicity (GRAVY)	-0.420

Table 2: Amino Acid Composition.

Amino Acid	No	Percentage
Ala (A)	22	11.8
Arg (R)	16	8.6
Asn (N)	3	1.6
Asp (D)	8	4.3
Cys (C)	0	0.0
Gln (Q)	8	4.3
Glu (E)	26	14.0
Gly (G)	8	4.3
His (H)	8	4.3
Ile (I)	8	4.3
Leu (L)	19	10.2
Lys (K)	9	4.8
Met (M)	7	3.8
Phe (F)	6	3.2
Pro (P)	4	2.2
Ser (S)	10	5.4
Thr (T)	10	5.4
Trp (W)	0	0.0
Tyr (Y)	0	0.0
Val (V)	14	7.5
Pyl (O)	0	0.0
Sec (U)	0	0.0

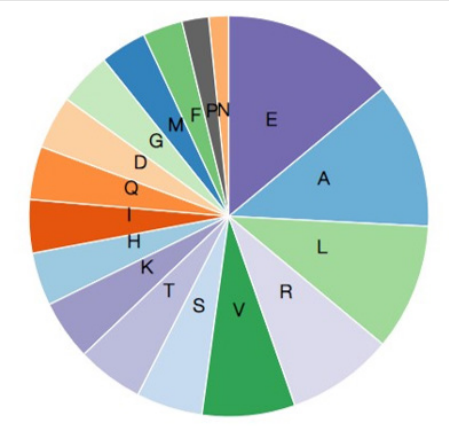

Figure 1: Amino Acid composition of VDM89547.1 protein.

Secondary Structure Prediction

The secondary structure prediction reveals important structural features of the 186 amino acid protein analyzed. SOPMA, using its default parameters, predicts a high percentage (74.73%) of residues forming alpha helices (Table 3). This suggests the protein likely contains several long helical regions that are stabilizing structural elements. In contrast, SOPMA predicts very little beta-sheet content (0% extended strand) (Figure 2). This indicates the protein does not contain extensive beta sheets, which often form protein aggregates. The prediction of 23.66% random coils suggests there are likely some flexible loop regions connecting the structured alpha helices. These loops may allow dynamic movements of the protein domains and could be important functional elements.

Table 3: Secondary Structure Elements.

Secondary Structure Elements	Values	Percentage
Alpha helix (Hh)	139	74.73%
3_{10} helix (Gg)	0	0.00%
Pi helix (Ii)	0	0.00%
Beta bridge (Bb)	0	0.00%
Extended strand (Ee)	0	0.00%
Beta turn (Tt)	3	1.61%
Bend region (Ss)	0	0.00%
Random coil (Cc)	44	23.66%
Ambiguous states	0	0.00%
Other states	0	0.00%

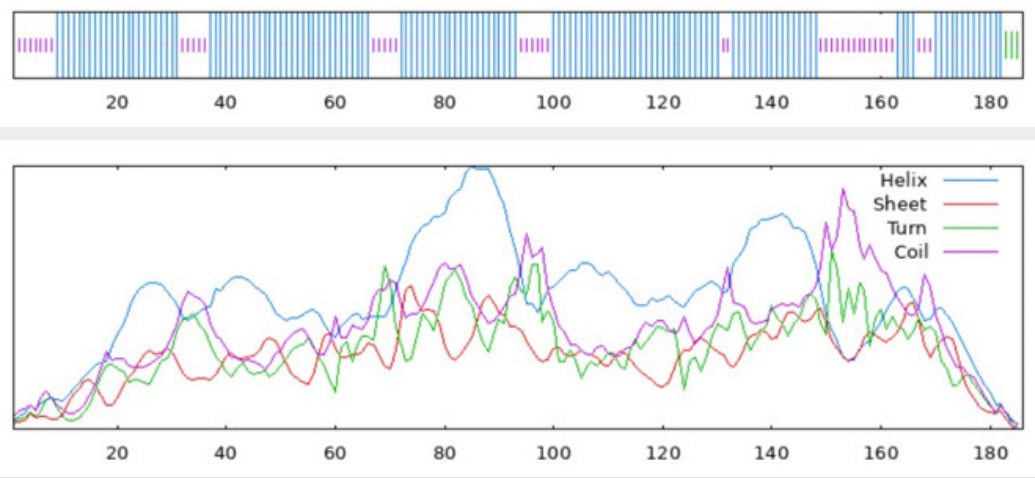

Figure 2: Secondary Structure of VDM89547.1 protein predicted by SOPMA.

The 1.61% beta turns also point to the existence of tight turns between alpha helices. The PSIPRED analysis provides additional confidence, with good agreement on the locations of helical and coil/loop sections. The multiple long helix regions are again evident. Collectively, the predictions paint a picture of a predominantly alpha-helical protein, with little beta-sheet content. The loop regions likely impart flexibility to allow functional conformational changes between structured helical domains. Further characterization is warranted to explore the functional implications of these structural elements. Analyzing the connectivity and length of specific helices, along with the sequence conservation of loop regions, could provide more insights.

Protein Binding Sites Prediction

The prediction of 17 distinct protein binding sites by the Predict Protein server provides insights into the likely functional regions of the protein. Binding sites allow a protein to interact with other molecules, enabling critical biological activities. The multiple binding sites identified suggest this protein is involved in several key molecular interactions. Specifically, the server pinpointed binding pockets of varying lengths across the protein sequence. The longest spans from residues 173-179, 153-156, and 58-62. These extensive sites indicate large interaction interfaces, which may be important for stable protein-protein or protein-ligand binding. Other shorter sites, like residues 180-181 and 171-172, likely represent smaller pockets that bind smaller compounds or contribute localized contacts for complex formation. Several binding pockets cluster together in the sequence, such as those spanning residues 113-124. This points to a possible functional hot spot, where multiple sites coordinate to create one larger interaction interface. The spacing of other sites suggests they mediate distinct functions. Analyzing the chemical properties and conserved residues of each predicted site could clarify the types of potential binding partners. The 17 binding sites highlight multiple regions primed for molecular interactions. This suggests the protein acts as a key node within a cellular network, binding various partners to coordinate multiple signaling pathways or structural complexes. Elucidating which proteins or other molecules occupy these sites will shed light on this protein's physiological role.

Tertiary Structure Prediction and Refinement

Predicting an accurate tertiary structure is critical for deducing the functional mechanisms of the target protein. We utilized a multi-step approach, first predicting an initial structure with AlphaFold 2, then refining it using GalaxyWeb (Figure 3). AlphaFold provided a reasonable first draft, while GalaxyWeb optimization enhanced the atomic positions and geometries. Assessment with MolProbity showed improvement in nearly all geometric criteria after refinement. Poor rotamers dropped from 1.28% to 0%, indicating better sidechain conformations. The Ramachandran outliers fell from 1.09% to 0%, with 100% of residues now in favored regions – this confirms refinement properly oriented the protein backbone. The clash score rose slightly from 2.38 to 4.76, potentially suggesting some minor unrealistic atomic contacts. However, the score remains low at 4.76 clashes per 1000 atoms. Overall, the MolProbity score improved from 1.47 to 1.25, validating enhanced model quality.

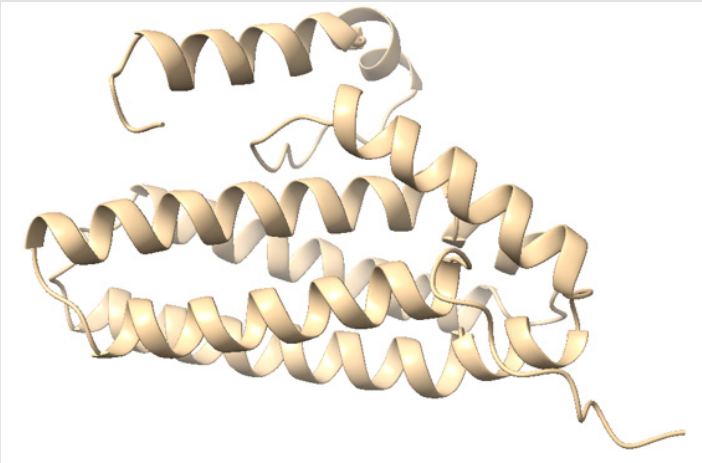

Figure 3: Visual representation of Tertiary Structure predicted by AlphaFold 2 and refined by GalaxyWeb Server.

Beyond geometry, the GalaxyWeb refinement boosted the ERRAT overall quality factor from 99.401 to 99.419, demonstrating improvement in non-bonded atomic interactions. Altogether, these metrics confirm that refinement corrected defects, producing a superior final model for analyzing structure-function relationships. Future work could involve molecular dynamics simulations to allow flexible movements and further refine the structure. Comparing the predicted model to experimentally determined structures of similar proteins

may also reveal areas for improvement. But fundamentally, the current refined model provides a high-quality template for elucidating the functional mechanisms of this important protein.

Tertiary Structure Validation

Validating the accuracy of the predicted tertiary structure is an essential final step before analyzing the functional implications. We utilized several complementary methods to rigorously assess model quality. The PROCHECK Ramachandran plot confirms that 98.8% of residues lie in the most favored regions, while 1.2% fall in allowed regions (Figure 4). The complete absence of outliers highlights excellent backbone geometries. Furthermore, all non-glycine/proline residues satisfy Ramachandran criteria, reflecting proper phi/psi angles. MolProbity also demonstrates favorable metrics – the score of 1.25 indicates only minor local backbone/sidechain defects. A 100% Ramachandran favored percentage reinforces that backbone conformations are physically realistic. The clash score of 4.76 also points to minimal issues with atomic contacts. The QMEAN Z-score provides a model quality rating by comparing structural features to experimental structures (Figure 5). Our protein received a high score of -5.84, confirming its similarity to native proteins. Collectively, these validation metrics signify the tertiary structure closely resembles the native state. This provides confidence in utilizing the model to infer functional mechanisms, interactions, and biological activities. Ongoing refinements via molecular dynamics simulations could further enhance accuracy. But overall the current structure represents a robust framework for deducing the protein's physiological role through detailed structural analysis.

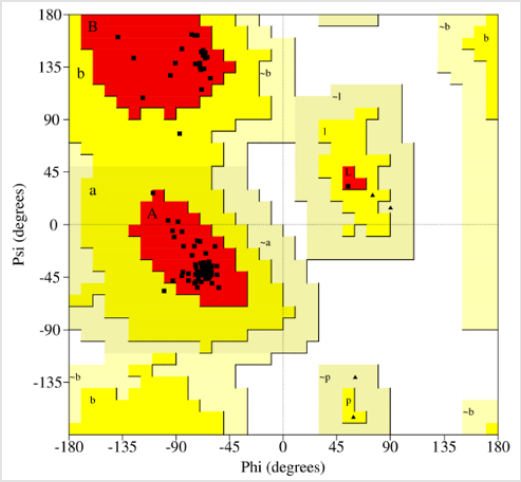

Figure 4: Ramachandran plot of the refined 3D structure of protein.

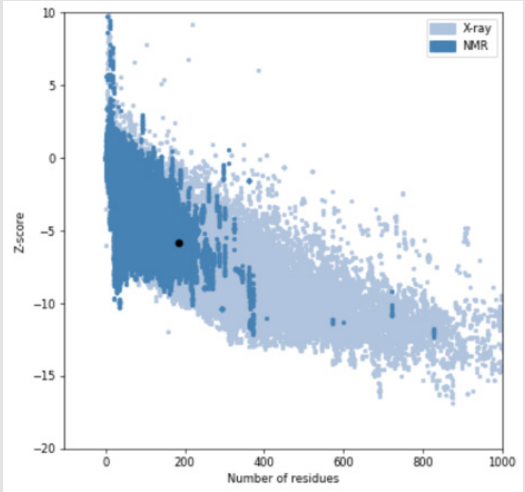

Figure 5: Z-score of refined tertiary structure of protein obtained from ProSA Webserver.

Molecular Docking

During the analysis, ClusPro was used to examine how the VDM89547.1 protein of *M. basiliense* interacts with human Glutathione (GSH). Upon a detailed examination of the docked conformations within this cluster, consistent and favorable binding patterns were observed between the VDM89547.1 protein and GSH (Table 4). The representative center-weighted score of -723.0 indicates a strong affinity, suggesting a stable binding interaction. Furthermore, the representative lowest energy-weighted score of -749.9 reinforces the robustness of this molecular association. The cluster has a substantial membership of 144, highlighting the consistency of these results across a diverse range of conformations. This adds to the reliability and significance of the observed interaction between VDM89547.1 protein and GSH (Figure 6).

Table 4: Balanced coefficient scores of docked complex of M. basiliense protein and GSH.

Cluster	Members	Representative	Weighted Score
	144	Center	-723.0
		Lowest Energy	-749.9

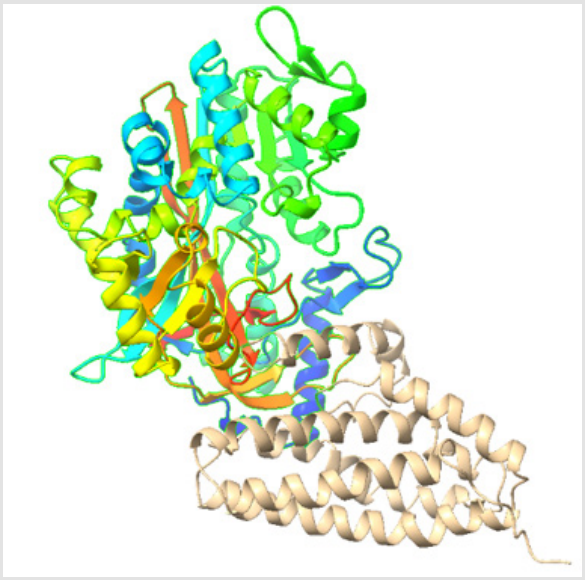

Figure 6: Docked complex of HGS and *M. basiliense* protein. The protein in color is HGS and the protein in blue is *M. basiliense* protein.

DMIM Plot

The DMIM (Distance and Environment Dependent Model Index Matrix) plot provides valuable insights into the interaction interface between the VDM89547.1 protein and glutathione (Figure 7). Specifically, it reveals the individual residue contacts mediating binding. Several VDM89547.1 residues form key interactions - Gln38, Arg142, Arg154, and Arg66 bind glutathione through predicted hydrogen bonds/salt bridges based on their charged natures (Table 5). Surrounding hydrophobic residues like Ile182, Leu163, and Ala157 likely contribute to van der Waals contacts. Notably, many of the glutathione residues engaged in these predicted contacts are also charged (Arg/Glu/Asp), indicating potential electrostatic interactions complement the hydrogen bonds. The multiple contacts distributed across the interface suggest an extensive binding surface, reinforcing the stability inferred from the favorable docking scores. Ultimately, the DMIM plot-identified interface provides a starting point for an atomic-level understanding of how the VDM89547.1 protein recognizes glutathione. This molecular recognition forms the foundation of their coupled functional roles in the cell - whether glutathione metabolism, redox regulation, or protective antioxidant pathways. Elucidating the structural determinants enabling this key protein-protein interaction remains an essential goal.

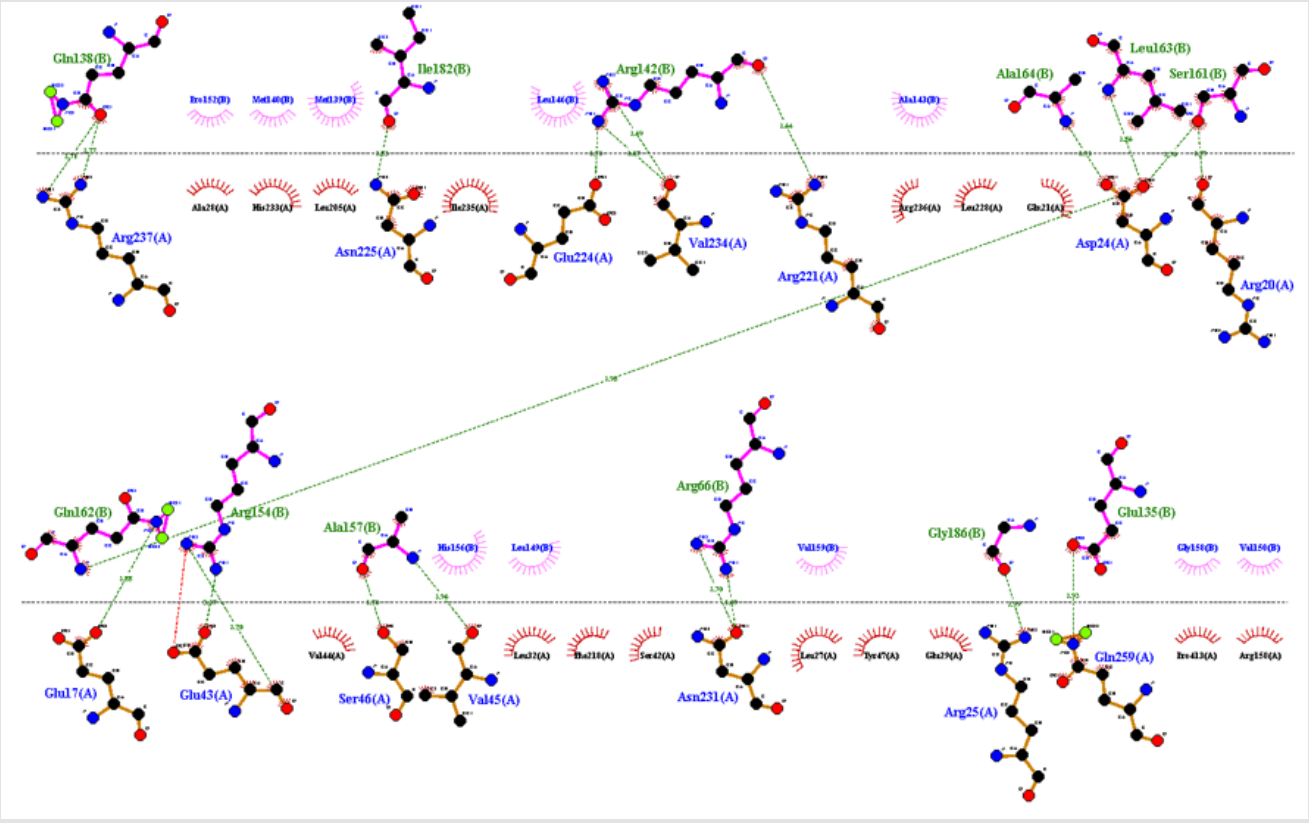

Figure 7: DMIM plot generated by LigPlot+.

Table 5: MolProbity score of the initial and refined model of M. basiliense protein.

		Initial Model		Refined Model	
All Atom-Contacts	**Clash score, all Atoms**	**2.38**		**4.76**	
		Clash score is the number of serious steric overlaps (>0.4A) per 1000 atoms.			
All Geometry	Poor Rotamers	2	1.28%	0	0.00%
	Favored Rotamers	151	96.79%	155	99.36%
	Ramachandran outliers	2	1.09%	0	0.00%
	Ramachandran favored	174	94.57%	184	100.00%
	Rama Distribution Z-score	0.66 ±0.61		1.73 ±0.53	
	MolProbity Score	1.47		1.25	
	Overall Quality Factor (ERRAT	99.401		99.419	

MD Simulation

The molecular dynamics (MD) simulation suggests that the protein maintains a stable overall structure in the aqueous environment (Figure 8). The Radius of gyration (Rg) plot demonstrated the increasing compactness of the protein structure in time. The values vary from 1.64 to 1.72 nm and, in general, exhibit a decreasing tendency. This implies that the protein has a fairly fixed global structure over the course of the simulation with a little compaction. The RMSD (Root Mean Square Deviation) plot describes the mean distance of the protein backbone from its starting conformation. It rises sharply in the initial 10 ns and later remains almost constant at the range of 0.4 – 0.5 nm. This suggest that the protein first experiences conformational changes and then stabilizes to a more fixed conformation. The plateau indicates that the simulation has probably achieved steadiness. The count of hydrogen bonds oscillates around 135–175 during the simulation. This variability is quite acceptable and arises from the fact that

hydrogen bonding in proteins is constantly changing. The presence of a large number of hydrogen bonds suggests that the protein retains its secondary and tertiary conformation. The total Solvent Accessible Surface Area (SASA) total becomes slightly smaller over time ranging from 115 nm^2 to 100 nm^2. This reduction in SASA is in agreement with the marginal compaction noted previously in the Rg plot. It indicates that the protein could be folding to a slightly more compact state perhaps by hydrophobic collapse or by changing its conformation. RMS fluctuation plot demonstrates the motional freedom of various residues in the protein. The majority of residues exhibit small fluctuations (below 0.5 nm), which suggests a stable conformation. But there is high density at the start and the end of the sequence, which probably represent more mobile N- and C-terminal tails.

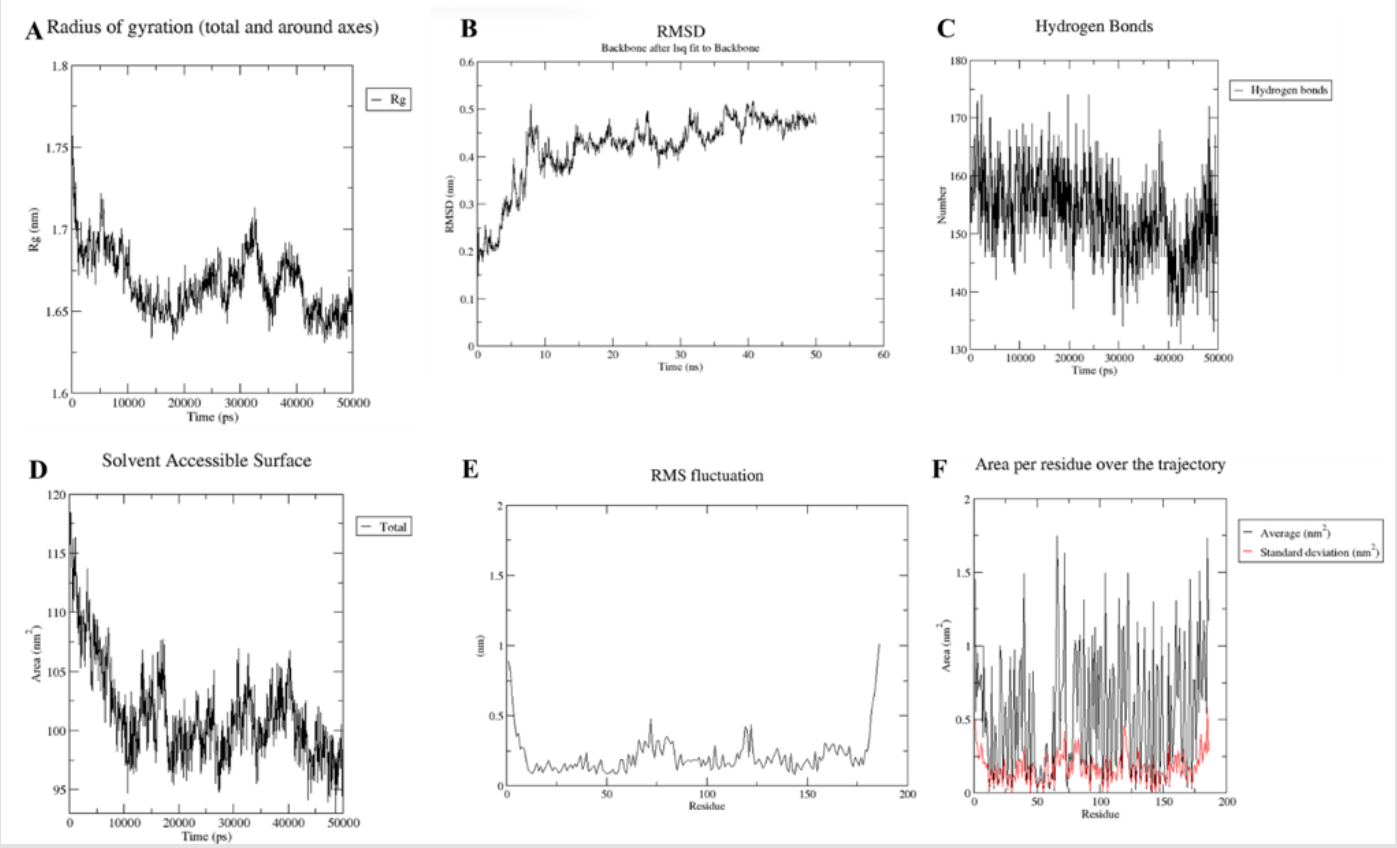

Figure 8: Molecular dynamics simulation analysis of protein structure and dynamics in aqueous solution.

A. Radius of gyration (Rg) over 50 ns, indicating overall protein compactness.

B. Root Mean Square Deviation (RMSD) of backbone atoms over 60 ns, showing structural stability after initial equilibration.

C. Number of intramolecular hydrogen bonds over 50 ns, demonstrating maintenance of secondary and tertiary structure.

D. Total Solvent Accessible Surface Area (SASA) over 50 ns, revealing slight compaction.

E. Root Mean Square (RMS) fluctuation per residue, highlighting regions of structural flexibility.

F. Average solvent-exposed area per residue (black) and its standard deviation (red) over the trajectory, providing insights into residue-level solvent accessibility and dynamics. These plots collectively characterize the protein's conformational behavior and stability throughout the simulation.

A few of these internal peaks may correspond to the loop regions with higher mobility. Area per residue over the trajectory graph represents the mean solvent accessible surface area per residue (black) and the standard error (red). It appears that residues with higher values are more solvent exposed while residues with lower values are more buried. The standard deviation of the exposure level indicates variability, which is informative of the dynamics of different protein regions. Consequently, these outcomes indicate that protein retains a fairly constant global conformation throughout the simulation, although it does undergo a slight compaction. The initial rise in RMSD after which the value stabilizes suggests that the system has probably achieved equilibrium. The fact that hydrogen bonds remain stable and the patterns in SASA and per-residue area suggest the stability and movement of the protein.

Conclusion

This study presented a comprehensive in silico analysis of the Hemerythrin-like protein VDM89547.1 from the emerging pathogenic bacterium *Mycobacterium basiliense*. We utilized a multitude of computational approaches to predict key structural features and explore the functional implications of this uncharacterized protein. Physicochemical calculations revealed several distinctive properties of VDM89547.1, including thermolability, high aliphatic index, acidic pI, and hydrophilicity. Secondary structure predictions painted a picture of a highly helical protein likely containing flexible loop regions. Tertiary modeling and refinement generated an accurate 3D structural model, with validation indicating similarity to native protein folds. Molecular docking simulations predicted favorable binding between VDM89547.1 and glutathione. The extensive interface was mediated through putative hydrogen bonds and salt bridges.

Molecular dynamics simulations in aqueous solvent confirmed overall structural stability during dynamics. Collectively, these analyses provide the first insights into the structure-function relationships of VDM89547.1. The inherent instability and multitude of predicted binding pockets suggest this protein readily engages in transient heat-sensitive interactions integral to *M. basiliense* survival or pathogenicity. Elucidating specific binding partners and clarifying VDM89547.1's physiological role remains imperative. As molecular mechanisms of emerging mycobacterial pathogens are scarcely understood, this work lays the foundation for future experiments probing this novel bacterial protein. Ultimately, comprehensively characterizing such unexplored proteins could reveal new diagnostic markers or therapeutic targets against *M. basiliense* infection.

Future Perspectives

This *in silico* investigation of the unexplored VDM89547.1 protein sets the stage for numerous avenues of future research to elucidate its structure-function relationships and role within *M. basiliense*. Experimentally determining the tertiary structure of VDM89547.1 via X-ray crystallography or NMR spectroscopy is imperative. This would enable unambiguous validation of the computationally derived model, provide atomic-resolution details of the distinctive binding pockets, and reveal insights into dynamics and conformational changes. Further computational analyses could also be undertaken. Conducting microsecond-scale molecular dynamics simulations may reveal additional dynamic structural transitions relevant to function. Ligand screening and virtual fragment-based drug design could identify compounds that favorably bind VDM89547.1, informing future laboratory assays. Sequence analyses across mycobacterial strains could highlight conserved residues that define fundamental functions. Most importantly, the biological activity and native binding partners of VDM89547.1 remain unknown.

Affinity purification experiments could elucidate its interaction network within *M. basiliense*, shedding light on associated pathways and processes. Phenotypic screens using gene knockout/overexpression strains would clarify how modulating VDM89547.1 levels impacts *M. basiliense* physiology and pathogenesis. Structural mapping of binding interfaces would pinpoint molecular determinants governing interactions. Overall, this work sets the scaffold for the multifaceted characterization of an important yet remarkably unexplored bacterial protein. Ultimately, exploring avenues from detailed structural analyses to unraveling functional implications within the cell will provide fundamental, translationally valuable insights into emerging mycobacterial pathogens. The path forward will integrate computational, biochemical, and microbiological techniques to unravel mysteries that could transform infection treatment paradigms.

Declarations

Acknowledgements

Not Applicable.

Funding

Not Applicable.

Competing Interests

Authors do not have any kind of competing interests.

References

1. Seth Smith H, Frank Imkamp, Florian Tagini, Aline Cuénod, Rico Hömke, et al. (2019) Discovery and characterization of Mycobacterium basiliense sp. nov., a nontuberculous Mycobacterium isolated from human lungs. Frontiers in microbiology 9: 3184.
2. Donohue MJ, L Wymer (2016) Increasing prevalence rate of nontuberculous mycobacteria infections in five states, 2008–2013. Annals of the American Thoracic Society 13(12): 2143-2150.
3. Guarner J (2018) Buruli ulcer: review of a neglected skin mycobacterial disease. Journal of clinical microbiology 56(4): e01507-17.
4. Armstrong DT, N Parrish (2021) Current updates on mycobacterial taxonomy, 2018 to 2019. Journal of clinical microbiology 59(7): e01528-20.
5. Jenuth JP (1999) The NCBI: Publicly available tools and resources on the web. Bioinformatics methods and protocols 132: 301-312.
6. Gasteiger E, Christine Hoogland, Alexandre Gattiker, Severine Duvaud, Marc R. Wilkins, et al. (2005) Protein identification and analysis tools on the ExPASy server.
7. Combet C, C Blanche, C Geourjon, G Deléage (2000) NPS@: network protein sequence analysis. Trends in biochemical sciences 25(3): 147-150.
8. Buchan DW, DT Jones (2019) The PSIPRED protein analysis workbench: 20 years on. Nucleic acids research 47(W1): W402-W407.
9. Ko J, Hahnbeom Park, Lim He, Chaok Seok (2012) GalaxyWEB server for protein structure prediction and refinement. Nucleic acids research 40(W1): W294-W297.

10. Laskowski RA, JM Thornton, DS Moss, MW MacArthur (1993) PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography 26(2): 283-291.

11. Williams CJ, Jeffrey J Headd, Nigel W Moriarty, Michael G Prisant, David C Richardson, et al. (2018) MolProbity: More and better reference data for improved all-atom structure validation. Protein Science 27(1): 293-315.

12. Wiederstein M , MJ Sippl (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.

13. Kozakov D, David R Hall, Bing Xia, Kathryn A Porter, Dzmitry Padhorn et al. (2017) The ClusPro web server for protein–protein docking. Nature protocols 12(2): 255-278.

14. Pettersen EF, Thomas D Goddard, Conrad C Huang, Elaine C Meng, Gregory S Couch, et al. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Science 30(1): 70-82.

15. Bekker H, hjc berendsen, ej dijkstra, s achterop, r vondrumen, et al. (1993) Gromacs-a parallel computer for molecular-dynamics simulations. in 4th international conference on computational physics (PC 92).

16. Waterhouse A, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, et al. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic acids research 46(W1): W296-W303.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2025.61.009581

Mubeen Ali Niaz. Biomed J Sci & Tech Res

Submission Link: https://biomedres.us/submit-manuscript.php