Original Research Article

Structural analysis and functional annotation of a hypothetical protein from *Streptococcus mitis, a*viridans group streptococci (VGS) *using an in-silico*approaches

ABSTRACT

Streptococcus mitis, a member of the viridansgroup streptococci (VGS), is commonly foundin the normal flora of the human oropharynx. and can cause various infections in humans, including orbital cellulitis, infective endocarditis, and bacteremia in neutropenic individuals. Identifying specific proteins, particularly hypothetical proteins (HPs), is crucial for developing effective treatments. This study focused on characterizing an unannotated HP from *Streptococcus mitis* (accession no. BCJ11593.1), predicting its properties and structure using computational tools, and validating these predictions. The HP was identified as a cytoplasmic and stable protein. NCBI-CD Search, and InterProScan, two functional annotation tools predicted that target HP was ribosome-associated protein YbcJ, specifically the S4-like RNA binding protein. Secondary structural investigations found that the alpha helix was prevalent.The study determined its three-dimensional (3D) structure through homology modeling by the SWISS-MODEL server and verified using quality evaluation tools such as PROCHECK, QMEAN, ERRAT and ProSA.The findings lay the groundwork for potential antibacterial treatments, highlighting the importance of understanding the structure and function of specific proteins in *Streptococcus mitis*.

Keywords: *Functional annotation, Hypothetical protein, In silico characterization, Streptococcus mitis, Three-dimensional structure.*

1. INTRODUCTION

Next-generation sequencing (NGS) allows researchers to quicklyaccumulate enormous volumes of data. However, as more organisms undergo sequencing, assigning genetic functions becomes increasingly challenging[1-2].The term "Hypothetical Proteins (HPs)" refers to proteins with

unknown molecular functions, constituting over 30% of proteins in many species [3]. *In-silico* characterization of HPs facilitates the determination of three-dimensional (3D) structures, unveiling new domains, motifs, pathways, protein networks, and other valuable information [4-6]. Structural and functional annotation of HPs not only establishes a link to potential therapeutic targets and biomarkers but also aids in understanding their roles [7]. To effectively annotate the roles of HPs in various disease-causing microorganisms, several bioinformatics databases and methodologies have been employed[8-11].

Streptococcus mitis, a Gram-positive coccus in the Viridians Group Streptococci (VGS) and mitis groups, functions as a commensal bacterium in the oral cavity by adhering to dental hard tissues and mucous membranes[12-14]. It is widely present in the in the normal skin, integumentary system, gastrointestinal tract, female vaginal tract, and oropharynx flora[13, 15-17].However, *Streptococcus mitis*is linked to diverse diseases and infections within the human body, functioning as an opportunistic pathogen in immunocompromised individuals.VGS has the potential to trigger invasive conditions, including bloodstream infections, pneumonia, endocarditis, enteritis, and meningitis, particularly in patients with compromised immune status or other risk factors[16, 18-20]. Furthermore, serious clinical diseases such VGS shock syndrome are caused by *Streptococcus mitis* strains in cancer patients[18].Maternal septicemia and intra-amniotic infection resulting from *Streptococcus mitis*, potentially linked to periodontitis, may be observed in women experiencing preterm premature rupture of membranes (PROM) [21]. The emergence of viridansstreptococci, notably *Streptococcus mitis*, exhibiting high resistance to penicillin and leading to sepsis and meningitis in patients with leukemia, lymphoma, or neutropenia, is a matter of considerable concern[14, 22].

An in-silico analysis of a potential protein from *Streptococcus mitis* is essential because understanding the genome of this bacterium might help develop useful medications or vaccinations. effective drugs or vaccines. In this study, a HP from *Streptococcus mitis* with accession number BCJ11593.1 was selected and will undergo in-depth structural and functional analysis utilizing a variety of bioinformatics tools.All of the tools and programs used for the functional annotation of *Streptococcus mitis* HPs are listed in Table 1.

Table 1.Tools used for the in-silico analysis of the hypothetical protein

Function	Tools/Server	URL
Sequence retrieval	NCBI	https://www.ncbi.nlm.nih.gov/
Analysis of Physiochemical properties	ExPASyProtParam	https://web.expasy.org/protparam/
Subcellular localization	CELLO	http://cello.life.nctu.edu.tw/
	PSORTb	https://www.psort.org/psortb/
Functional characterization	Conserved Domain Database	https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
	InterProScan	http://www.ebi.ac.uk/Tools/services/web/toolform.ebi?tool=iprscan5
Sequence similarity search	BLASTp	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Secondary structure prediction	PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred/
	SOPMA	https://npsaprabi.ibcp.fr/cgibin/npsa_automat.pl?page=/NPSA/npsa_sopma.html
3D Structure prediction	SWISS-MODEL	http://swissmodel.expasy.org/
3D Structure validation	PROCHECK, Verify3D & ERRAT	https://saves.mbi.ucla.edu/
	QMEAN4	https://swissmodel.expasy.org/qmean/
	ProSA-web (Z scores)	https://prosa.services.came.sbg.ac.at/prosa.php

2. MATERIALS AND METHODS

2.1 Sequence retrieval

The amino acid sequence of HP (accession no. BCJ11593.1) from *Streptococcus mitis* was obtained in FASTA format from the protein database of the National Center for Biotechnology Information (NCBI) [23]. In the virtual annotation process, the protein sequence was then submitted to various prediction servers.

The NCBI-provided FASTA format is:

>BCJ11593.1 hypothetical protein SMNM65_20250 [*Streptococcus mitis*]

MEYKLFEEFITLQALLKELGITHSGGAIKSFLSEHSVYFNGELESRRGKKLRIGDKVDIPD
MNIDILLTQPTSEEQDEYQADKVEKERIAKLVKEMNKGVKKDKSKPTSSPKSKQAPRFP
GR

2.2 Analysis of physicochemical properties

The ExPASyProtParam tool [24]was utilized to investigate the physicochemical attributes of HP. This analysis encompassed parameters such as molecular weight, theoretical isoelectric point (pI), amino acid composition, total count of positive and negative residues, instability index, aliphatic index (AI), grand average of hydropathicity (GRAVY), molecular formula, and estimated half-life of the protein.

2.3 Subcellular localization prediction

Subcellular localization is crucial for understanding the function of proteins and for the study of the genome. Both CELLO [25]and PSORTb3.0 [26]estimated the HP subcellular position.

2.4 Functional annotation

Proteins, intricate molecules with diverse roles in living organisms, are categorized into various families and superfamilies based on shared sequence characteristics, domains, motifs, and functional attributes [27]. The NCBI Conserved Domain Search Service (CD Search) [28] and InterProScan[29]were employed to discern the potential function and identify conserved domains of the target protein.

2.5 Multiple sequences alignment and phylogenetic tree analysis

To identify homologous proteins, a BLASTp search was performed on the NCBI website against the nonredundant database, utilizing the default settings [30]. Then, using CLC Sequence Viewer version 8, a multiple sequence alignment and a tree of phylogeny were produced. This software is widely employed for sequence analysis and visualization, facilitating the comparative analysis of protein sequences and the construction of phylogenetic relationships.

2.6 Secondary structure determination

The self-optimized prediction method with alignment, or SOPMA[31], was applied to predict the two-dimensional structure of the HP. PSIPRED [32]was employed as an additional method to confirm the results of SOPMA.

2.7 Homology modeling

Based on homology modeling, the target protein's three-dimensional (3D) structure was ascertained using the SWISS-MODEL [33]server. For every protein sequence, the server automatically performs a BLASTp search to find templates. With 96.72% sequence identity, template protein Q8DMX4.1.A was chosen for homology modeling based on the query result. The 3D model structure viewed with PyMOL v2.0.

2.8 Quality assessment

The assessment of the anticipated 3D structure's validity was conducted through the SAVES server modules PROCHECK [34] and ERRAT [35]. Additionally, the QMEAN programs [36] on the ExPASy server of the SWISS-MODEL Workspace were employed to determine the QMEAN Z-score and evaluate the model's quality. The ProSA-web server [37] was utilized to estimate Z scores for obtained model.

3. RESULTS

3.1.Analysis of physicochemical properties

The ProtParam program was employed to calculate various physicochemical characteristics of the hypothetical protein (accession no. BCJ11593.1), as detailed in Table 2. The protein was predicted to consist of 122 amino acids, with a grand average of hydropathicity (GRAVY) of -0.807, a molecular weight (M.W.) of 13,910.97, a theoretical isoelectric point (pI) of 9.05, and a total of 20 negatively charged residues (Asp + Glu) and 23 positively charged residues (Arg + Lys). The anticipated half-life (HL) of the target protein was 30 hours, and its instability index (II) was calculated to be 35.04, indicating stability. The aliphatic index (AI) was found to be 76.72, suggesting that the protein is likely stable across a broad range of temperatures.

Table 2. Physiochemical properties of HP estimated by ProtParam tool

Description	Value
Number of amino acids	122
Molecular weight (Da)	13910.97
Theoretical pI	9.05
Number of positively charged residues	23
Number of negatively chargedresidues	20
Instability index	35.04
Aliphatic index	76.72
Grand average of hydropathicity (GRAVY)	-0.807

3.2.Subcellular localization prediction

As various cellular compartments correlate to different functions, identifying the hypothetical protein's subcellular location is crucial to understanding its function. This information is valuable for potential drug targeting against the identified protein. According to predictions from CELLO, our target protein is anticipated to be cytoplasmic. Consistently, the PSORTb server also predicted the protein's subcellular location to be in the cytoplasm.

3.3.Protein family analysis

To unveil conserved domains and infer the potential function of our target protein, various annotation tools were employed. According to predictions from NCBI-CD Search, the target protein was identified to harbor the S4 superfamily domain and was categorized as a ribosome-associated protein YbcJ, specifically a S4-like RNA-binding protein. The S4 superfamily domain, predicted by the NCBI-CD server, spanned amino acid residues 1-122 with an E-value of 6.57e-32. Additionally, the InterProScan server concurred by predicting the presence of the RNA-binding S4 superfamily domain in the target protein.

3.4.Multiple sequence alignment and Phylogeny analysis

The BLASTp search against the non-redundant database revealed similarities between our target protein and other known S4 domain-containing proteins, such as YaaA from various bacteria (refer to Table 3). Multiple sequence alignments were conducted on a subset of proteins selected from the BLASTp results to identify conserved and distinctive residues among the homologs (Fig. 1). Based on the same dataset, a phylogenetic tree (Fig. 2) was constructed, indicating a shared evolutionary origin between the target protein and *Streptococcus* (WP_000455897.1).

Table 3.BLASTp result shows similarities between proteins

Accession no.	Organism	Protein name	Score	Percent Identity	E-value
WP_000455897.1	*Streptococcus*	S4 domain-containing protein YaaA	247	100.00%	2e-82
CIY83532.1	*Streptococcus pneumoniae*	S4 domain-containing protein YaaA	246	99.18%	4e-82
WP_049532442.1	*Streptococcus pseudopneumoniae*	S4 domain-containing protein YaaA	246	99.18%	4e-82
WP_247942388.1	*Streptococcus mitis*	S4 domain-containing protein YaaA	245	98.36%	2e-81
CON46485.1	*Streptococcus pneumoniae*	S4 domain-containing protein YaaA	244	98.36%	2e-81

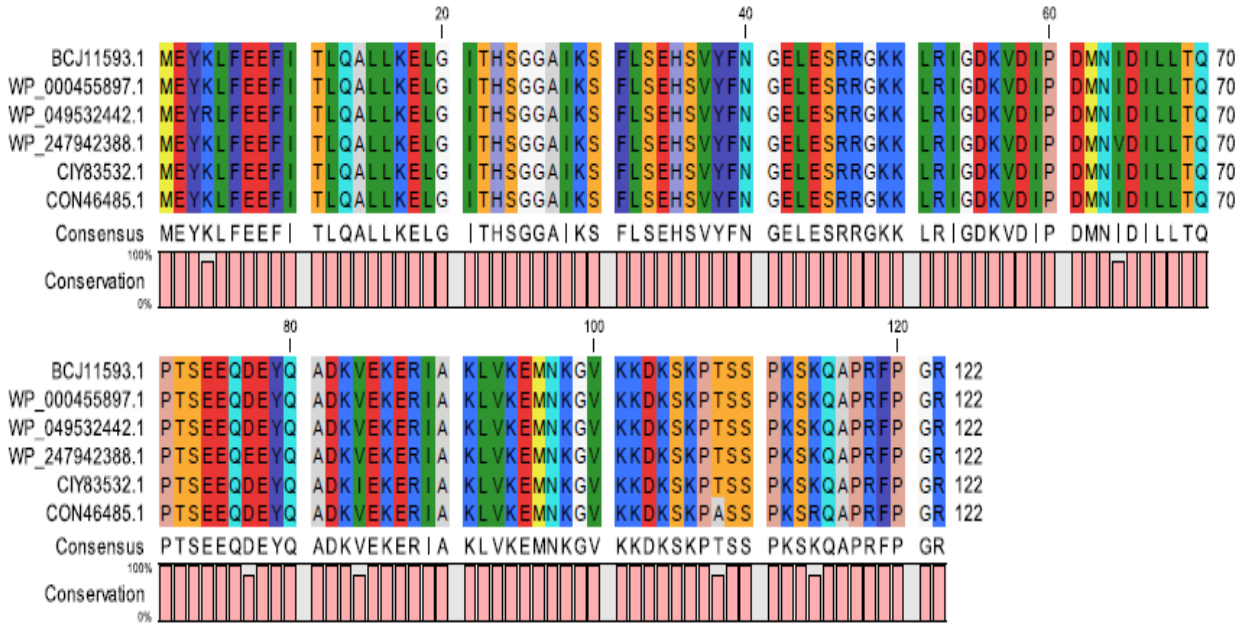

Fig. 1. Multiple sequence alignments among different S4 domain-containing protein YaaAusing CLC sequence viewer

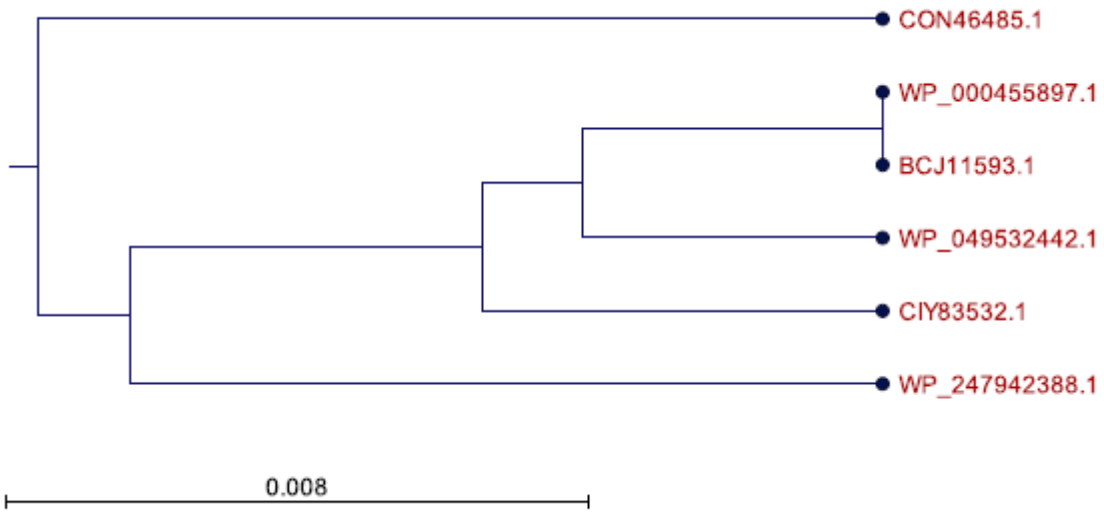

Fig. 2.A phylogenetic tree showing evolutionary relationship of the target protein with other S4 domain-containing protein,YaaA.

3.5. Secondary structure determination

The determination of the protein's secondary structure was carried out using PSIPRED and the SOPMA server. Among the identified secondary structures, the alpha helix was the most prominent, alongside the random coil, beta-turn, and extended strand. According to the SOPMA analysis (Fig. 3), the distribution of secondary structures was as follows: random coil accounted for 32.79%, alpha helix for 44.26%, beta-turn for 7.38%, and extended strand, in conjunction with the random coil, constituted 15.57%. This finding was corroborated by a similar outcome obtained from the PSIPRED server (Fig. 4).

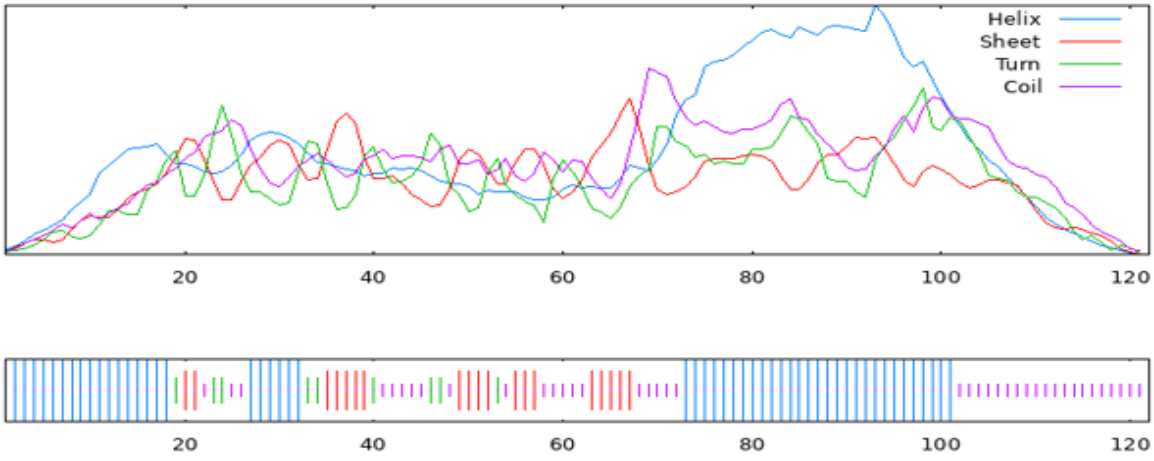

Fig.3. SOPMA server predicted the secondary structure

Fig. 4.PSIPRED predicted secondary structure of the target protein

3.6.Tertiary structure prediction

The target protein's tertiary structure was determined using template Q8DMX4.1.A from the SWISS-MODEL database, which has 96.72% sequence identity with the target protein. Figure 5 displays the three-dimensional structure of the protein using PyMOL 2.0.

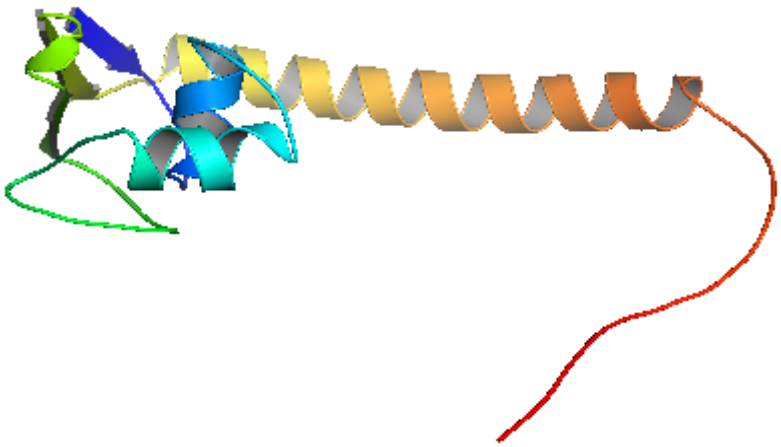

Fig. 5. Generated the target protein's three-dimensional structure using SWISS-MODEL (visualized using PyMOL program)

3.7.Model quality assessment

The assessment of the three-dimensional structure models was conducted utilizing PROCHECK, QMEAN, and ERRAT. The PROCHECK analysis revealed that 93.4% of amino acid residues occupied the most preferred zone in the "Ramachandran plot" (Table 4 and Fig. 6). With a commendable quality rating of 97.059, ERRAT indicated that the protein structure exhibited high quality (Fig. 7). The QMEAN tool placed the model within the permissible black zone with a QMEAN4 score of 0.50 (Fig. 8). To ascertain whether the input structure falls within the typical range of scores for native proteins of comparable size, the Z score, representing the overall model quality, was determined. The ProSA web server yielded a Z-score of -5.18 for the model target protein (Fig 9).

Table 4. Statistics of the target protein made by Ramachandran plot

Statistics	Number of AA residue	Percentage (%)
Residues in the most favored regions [A, B, L]	99	93.4%
Residues in the additional allowed regions [a, b, l, p]	6	5.7%
Residues in the generously allowed regions [~a, ~b, ~l, ~p]	1	0.9%
Residues in disallowed regions	0	0.00%
Number of non-glycine and non-proline residues	106	Total= 100%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residue	8	
Number of proline residues	6	
Total number of residues	122	

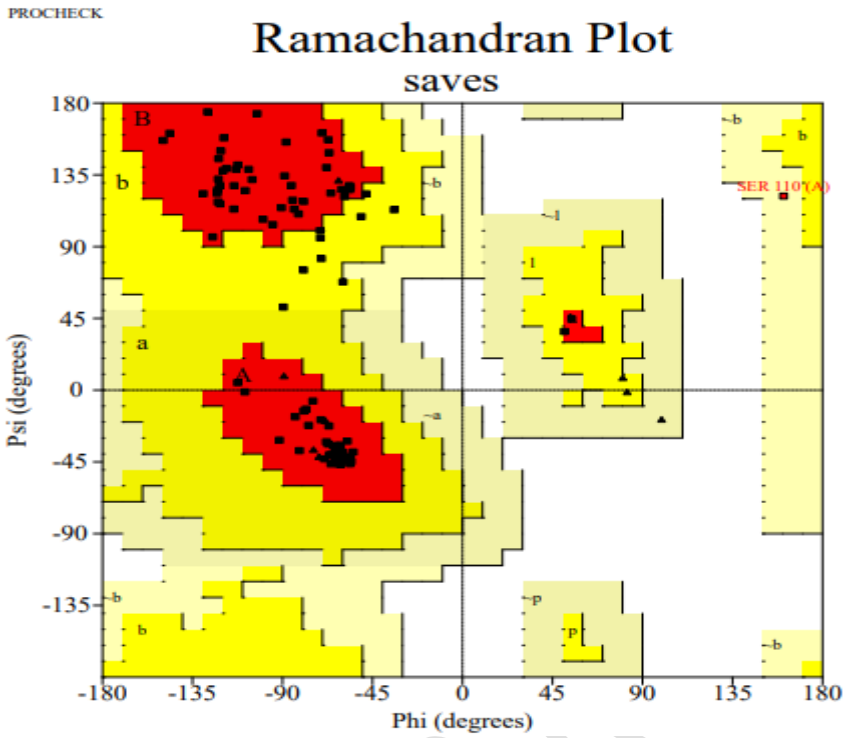

Fig.6. Ramachandran plot of model structure validated by PROCHECK program

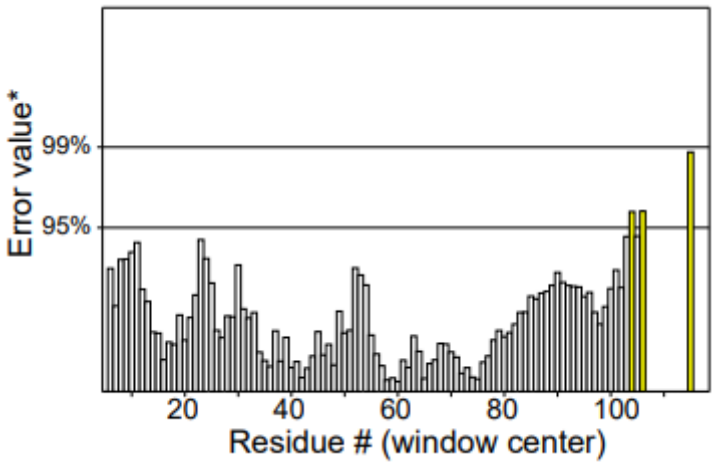

Fig.7. ERRAT output.Two lines on the error axis represent the confidence required to reject areas that exceed the error value

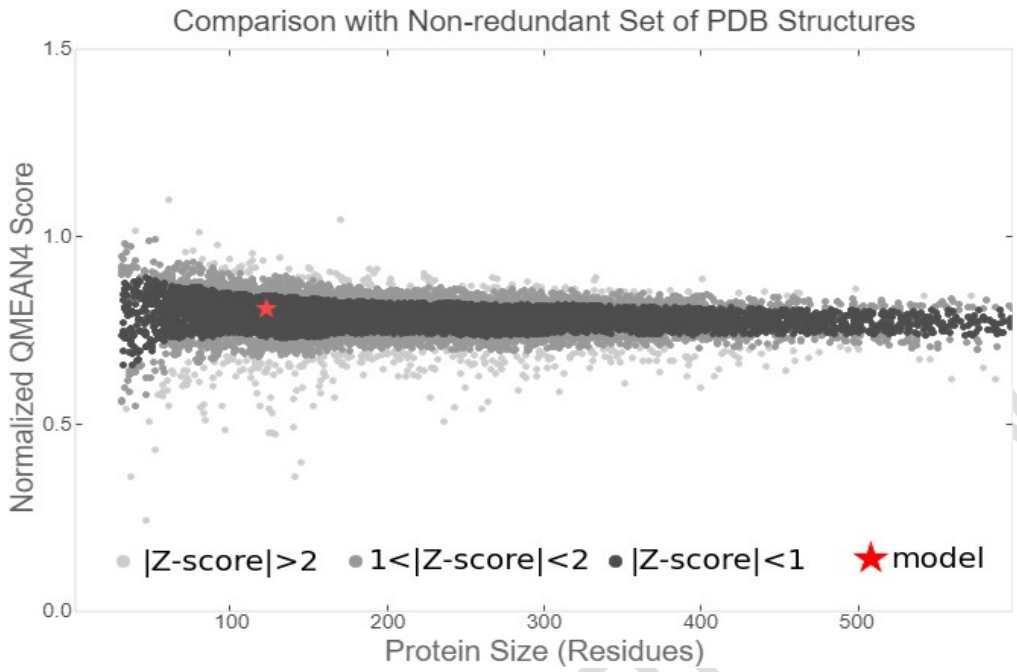

Fig.8. Graphical representation of QMEAN result of the model structure

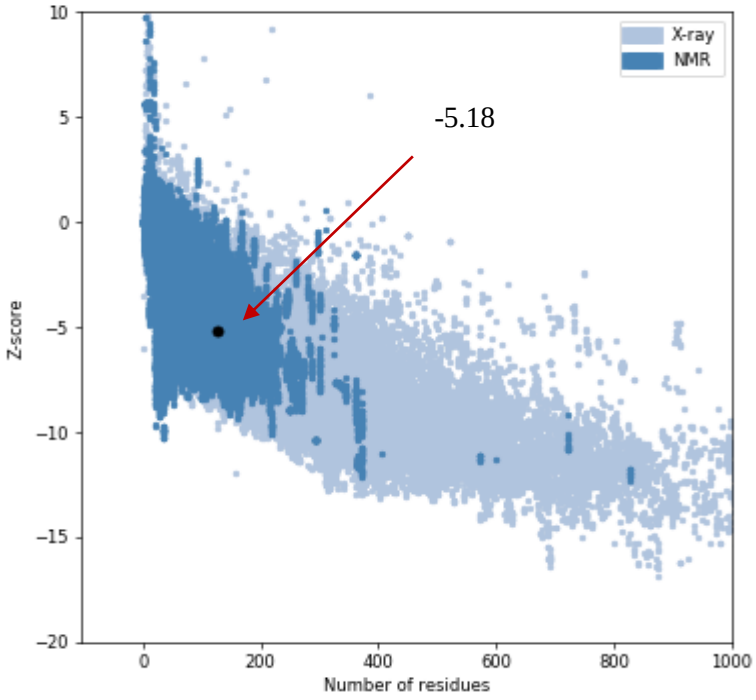

Fig. 9.Z score of the target protein using the ProSA server

4. DISCUSSION

Numerous ongoing studies focus on elucidating the structures and functions of hypothetical proteins. Some investigations delve into targeting these proteins for potential applications in various diseases[38]. The characterization of hypothetical proteins contributes to an enhanced understanding of bacterial metabolic pathways, aids in the development of pharmaceutical agents, and informs strategies for managing diseases [39]. Using a range of bioinformatics methods, the HP protein (accession no. BCJ11593.1) of the *Streptococcus mitis* strain was structurally and functionally described in this work. The protein is hydrophilic, or soluble in water, according to a grand average of hydropathicity (GRAVY) of -0.807, a theoretical pI of 9.05 that suggests the protein is acidic (pH > 7), and an instability index (II) of 35.04 (Table 2) shows that the protein is stable. This protein is expected to be located in the cytoplasm, according to the CELLO and PSORTb servers. By using domain and motif analysis, our target hypothetical protein was highly confidently predicted by all annotation techniques to be the ribosome-associated protein YbcJ, S4-like RNA binding protein.*Escherichia coli* protein YbcJ, belongs to the S4-like RNA binding domains, known for binding structured RNA in tRNA, rRNA, and mRNA pseudoknots[40-41]. As a 77-residue protein in a conserved bacterial family, YbcJ shares structural and sequence features with S4 superfamily RNA-binding proteins[42]. Its solution structure indicates the presence of an αL motif rich in positively charged residues (arginines and lysines), potentially involved in RNA binding (Volpon et al., 2003). The predominant feature in the protein's secondary structure is the alpha helix, which encompasses other components such as random coil, beta-turn, and extended strand. As indicated by the SOPMA analysis (Figure 3 and Figure 4), the distribution of secondary structures was as follows: random coil accounted for 15.57%, alpha helix for 44.26%, beta-turn for 7.38%, and extended strand, in conjunction with random coil, constituted 15.57%. The tertiary structure of the target protein was obtained from the SWISS-MODEL server, utilizing the template Q8DMX4.1.A, which exhibited a sequence similarity of 96.72% with the target protein (Figure 5). The three-dimensional (3D) structure created through the SWISS-MODEL server successfully passed all model quality assessment tools, including PROCHECK, QMEAN, and ERRAT. According to PROCHECK data (Table 4 and Figure 6), 93.4% of amino acid residues fell within the most preferred zone of the "Ramachandran plot." ERRAT predicted a high-quality protein structure with a quality factor of 97.059 (Figure 7). In comparison to other experimental

structures of a comparable size, the model also obtained a QMEAN4 score of 0.50, placing it in the desired dark grey zone (Figure 8). All of these findings point to the high quality of the predicted protein model, which makes it useful for more thorough research and analysis. The Z-score was determined using the ProSA web service, which offers a metric to assess the quality of the predicted protein model. The Z-score for the query model of the target protein was found to be -5.18 (Figure 9). This score is often used to assess if the input structure is within the score range seen for natural proteins of similar size. This Z-score suggests that the obtained model is reliable, and its overall quality is considered good. However, the research on hypothetical protein annotations helped in the creation of effective medications and vaccines to combat this particular sort of pathogen.

5. CONCLUSION

This study employed various bioinformatics methods to analyze a hypothetical protein sourced from *Streptococcus mitis*. The implications of our research extend to improving the functionality of the target protein and enhancing resource utilization efficiency. Future research will use structural and functional data to create new ligands for medication development and to experimentally test our findings. To improve future treatment approaches, it is imperative to continue investigating target proteins and their effectors in *Streptococcus mitis* and other species. This work contributes to understanding the structural and functional aspects of proteins with uncharacterized activities, providing a foundation for further research. The outcomes of this study may serve as a valuable resource for future in-silico investigations conducted by other researchers.

REFERENCES:

1. Choi HP, Juarez S, Ciordia S, Fernandez M, Bargiela R, Albar JP, Mazumdar V, Anton BP, Kasif S, Ferrer M, Steffen M. Biochemical characterization of hypothetical proteins from Helicobacter pylori. PLoS One. 2013 Jun 18;8(6): e66605.
2. Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. Genomics. 2008 Nov 1;92(5):255-64.

3. Shahbaaz M, Bisetty K, Ahmad F, Hassan I. Current advances in the identification and characterization of putative drug and vaccine targets in the bacterial genomes. Current topics in medicinal chemistry. 2016 Apr 1;16(9):1040-69.
4. Nimrod G, Schushan M, Steinberg DM, Ben-Tal N. Detection of functionally important regions in "hypothetical proteins" of known structure. Structure. 2008 Dec 12;16(12):1755-63.
5. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proceedings of the National Academy of Sciences. 1999 Apr 13;96(8):4285-8.
6. Idrees S, Nadeem S, Kanwal S, Ehsan B, Yousaf A, Nadeem S, Rajoka MI. In silico sequence analysis, homology modeling and function annotation of Ocimumbasilicum hypothetical protein G1CT28_OCIBA. International Journal Bioautomation. 2012;16(2):111.
7. Lubec G, Afjehi-Sadat L, Yang JW, John JP. Searching for hypothetical proteins: theory and practice based upon original data and literature. Progress in neurobiology. 2005 Sep 1;77(1-2):90-127.
8. Naqvi AA, Rahman S, Zeya F, Kumar K, Choudhary H, Jamal MS, Kim J, Hassan MI. Genome analysis of Chlamydia trachomatis for functional characterization of hypothetical proteins to discover novel drug targets. International journal of biological macromolecules. 2017 Mar 1; 96:234-40.
9. Naqvi AA, Anjum F, Khan FI, Islam A, Ahmad F, Hassan MI. Sequence analysis of hypothetical proteins from Helicobacter pylori 26695 to identify potential virulence factors. Genomics Inform. 2016 Sep 30;14(3):125-35.
10. Yang Z, Zeng X, Tsui SK. Investigating function roles of hypothetical proteins encoded by the Mycobacterium tuberculosis H37Rv genome. BMC genomics. 2019 Dec; 20:1-0.
11. Islam MS, Shahik SM, Sohel M, Patwary NI, Hasan MA. In silico structural and functional annotation of hypothetical proteins of Vibrio cholerae O139. Genomics & informatics. 2015 Jun;13(2):53.
12. Facklam R. What happened to the streptococci: overview of taxonomic and nomenclature changes? Clinical microbiology reviews. 2002 Oct;15(4):613-30.

13. Douglas CW, Heath J, Hampton KK, Preston FE. Identity of viridans streptococci isolated from cases of infective endocarditis. Journal of medical microbiology. 1993 Sep;39(3):179-82.
14. Jaing TH, Chiu CH, Hung IJ. Successful treatment of meningitis caused by highly-penicillin-resistant Streptococcus mitis in a leukemic child. Chang Gung medical journal. 2002 Mar 1;25(3):190-3.
15. Yiş R, Yüksel CN, Derundere U, Yiş U. Meningitis and white matter lesions due to Streptococcus mitis in a previously healthy child. MikrobiyolojiBulteni. 2011 Oct 1;45(4):741-5.
16. Kutlu SS, Sacar S, Cevahir N, Turgut H. Community-acquired Streptococcus mitis meningitis: a case report. International Journal of Infectious Diseases. 2008 Nov 1;12(6): e107-9.
17. Curtis H, Dirk G, Rob K, Sahar A, Badger JH, Chinwalla AT, Creasy HH, Earl AM, Fitzgerald MG, Fulton RS, Giglio MG. Structure, function and diversity of the healthy human microbiome. Nature. 2012; 486:207-14.
18. Shelburne SA, Sahasrabhojane P, Saldana M, Yao H, Su X, Horstmann N, Thompson E, Flores AR. Streptococcus mitis strains causing severe clinical disease in cancer patients. Emerging infectious diseases. 2014 May;20(5):762.
19. Fukayama H, Shoji K, Yoshida M, Iijima H, Maekawa T, Ishiguro A, Miyairi I. Bacterial meningitis due to the Streptococcus mitis group in children with cerebrospinal fluid leak. IDCases. 2022 Jan 1;27: e01406.
20. Tunkel AR, Sepkowitz KA. Infections caused by viridans streptococci in patients with neutropenia. Clinical infectious diseases. 2002 Jun 1;34(11):1524-9.
21. Chaemsaithong P, Lertrut W, Kamlungkuea T, Santanirand P, Singsaneh A, Jaovisidha A, Pakdeeto S, Mongkolsuk P, Pongchaikul P. Maternal septicemia caused by Streptococcus mitis: a possible link between intra-amniotic infection and periodontitis. Case report and literature review. BMC Infectious Diseases. 2022 Jun 20;22(1):562.
22. Balkundi DR, Murray DL, Patterson MJ, Gera R, Scott-Emuakpor A, Kulkarni R. Penicillin-resistant Streptococcus mitis as a cause of septicemia with meningitis in febrile neutropenic children. Journal of pediatric hematology/oncology. 1997 Jan 1;19(1):82-5.

23. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL. GenBank. Nucleic acids research. 2002 Jan 1;30(1):17-20.
24. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic acids research. 2003 Jul 1;31(13):3784-8.
25. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins: Structure, Function, and Bioinformatics. 2006 Aug 15;64(3):643-51.
26. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010 Jul 1;26(13):1608-15.
27. Wu CH, Huang H, Yeh LS, Barker WC. Protein family classification and functional annotation. Computational Biology and Chemistry. 2003 Feb 1;27(1):37-47.
28. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N. CDD/SPARCLE: the conserved domain database in 2020. Nucleic acids research. 2020 Jan 8;48(D1): D265-8.
29. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, Durbin R. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic acids research. 2001 Jan 1;29(1):37-40.
30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990 Oct 5;215(3):403-10.
31. Combet C, Blanchet C, Geourjon C, Deleage G. NPS@: network protein sequence analysis. Trends in biochemical sciences. 2000 Mar 1;25(3):147-50.
32. Buchan DW, Jones DT. The PSIPRED protein analysis workbench: 20 years on. Nucleic acids research. 2019 Jul 2;47(W1): W402-7.
33. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TA, Rempfer C, Bordoli L, Lepore R. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic acids research. 2018 Jul 2;46(W1): W296-303.

34. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography. 1993 Apr 1;26(2):283-91.
35. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein science. 1993 Sep;2(9):1511-9.
36. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics. 2011 Feb 1;27(3):343-50.
37. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research. 2007 Jul 1;35(suppl_2): W407-10.
38. Barragán-Osorio L, Giraldo G, J Almeciga-Diaz C, Aliev G, E Barreto G, Gonzalez J. Computational analysis and functional prediction of ubiquitin hypothetical protein: a possible target in Parkinson disease. Central Nervous System Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Central Nervous System Agents). 2016 Apr 1;16(1):4-11.
39. Sen T, Verma NK. Functional annotation and curation of hypothetical proteins present in a newly emerged serotype 1c of Shigella flexneri: emphasis on selecting targets for virulence and vaccine design studies. Genes. 2020 Mar 23;11(3):340.
40. Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature. 2000 Sep 21;407(6802):340-8.
41. Volpon L, Lievre C, Osborne MJ, Gandhi S, Iannuzzi P, Larocque R, Cygler M, Gehring K, Ekiel I. The solution structure of YbcJ from Escherichia coli reveals a recently discovered αL motif involved in RNA binding. Journal of bacteriology. 2003 Jul 15;185(14):4204-10.
42. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997 Sep 1;25(17):3389-402.