Genetics of Aquatic Organisms 9(2), GA884
https://doi.org/10.4194/GA884

PROOF
RESEARCH PAPER

GENAQUA
GENETICS OF AQUATIC ORGANISMS

In Silico Functional Annotation of E2102 Hypothetical Protein from *Aeromonas veronii*

Md Imtiaz Ahamed[1,*], Sadia Islam Mim[2], Most. Mahbuba Islam Emu[3]

[1]Jashore University of Science and Technology, Faculty of Biological Science, Department of Fisheries and Marine Bioscience, Jashore- 7408, Bangladesh.
[2]Govt. Michael Madhusudan College, Department of Zoology, Jashore, Bangladesh.
[3]Khulna University, Environmental Science Discipline, Khulna, Bangladesh.

How to Cite

Ahamed, M.I., Mim, S.I., Emu, M.M.I. (2025). In Silico Functional Annotation of E2102 Hypothetical Protein from *Aeromonas veronii*. *Genetics of Aquatic Organisms, 9(2), GA884*. https://doi.org/10.4194/GA884

Article History
Received 18 February 2025
Accepted 30 June 2025
First Online 22 July 2025

Corresponding Author
E-mail: imtiazahamed.85560@gmail.com

Keywords
Aeromonas veronii
Hypothetical protein
Functional annotation
Homology

Abstract

Many amino acids remain undiscovered and have not yet been associated with specific genome sequences. Modules of ambiguous activity are proteins detected through experimental methods that have not yet been assigned specific geometric annotations. In our study, the potential regulatory properties of the unidentified peptides from *Aeromonas veronii* (accession no. PXV57966.1) were explored and analyzed using several computational approaches and tools. From this in silico approach, the physical properties, cell position, 3-D arrangement, interactions among proteins, and functionality insights of the amino acid have all been determined. Protein-protein interactions were analyzed using STRING software, which revealed that E2102 protein interacts strongly with the peptidyl-tRNA hydrolase. The in-silico analysis indicated that the protein is hydrophilic, with its secondary structure mainly composed of alpha (α) helices. According to the result, the protein contains Ribosome-binding ATPase YchF domain, which suggests it may bind to ribosomal subunits. In addition, *Aeromonas veronii* is an opportunistic pathogen capable of causing various infections in humans, including gastroenteritis, soft tissue infections, and bacteremia. Thus, the study will contribute to the creation of novel antimicrobial treatments for treating severe intestinal tract infections by enhancing our understanding of the role of the E2102 domain.

Introduction

A significant portion of mammalian proteomes consists of hypothetical proteins, which are identified solely through nucleic acid sequence predictions and protein sequences with unknown functions (Lubec et al., 2005). Scientists have developed several techniques to predict protein function by employing various computational tools. This has been achieved through the utilization of information from similarities in sequence, the study of phylogenetics, protein-protein relationships, interactions among various proteins, the site of activity residue similarities, homologous categories, regulatory sites, and variation of gene activity. The conventional approach to determining function relies on similarities in sequences, utilizing tools like BLAST, FASTA, and PSI-BLAST (Bharat Siva Varma et al., 2015; Pearson, 2013). Hypothetical proteins are predicted proteins determined by sequences of nucleic acid, but they are not supported by research chemical data. Moreover, these proteins are distinguished by a low degree of identity to known, annotated proteins (Lubec et al., 2005). Some hypothetical proteins are common and can be identified in lifeforms from different evolutionary lineages. Hypothetical proteins account for a significant portion of the genes in the sequenced genomes of organisms. However, they have not been functionally characterized

or described at the protein chemical level (Galperin & Koonin, 2004; Reichart et al., 2020). Two types of hypothetical proteins exist. One group comprises uncharacterized protein families (UPFs), while the second includes domains with unknown functions (DUFs). Unidentified proteins have experimentally determined structures but remain uncharacterized and unlinked to any identified genes. DUFs are experimentally identified proteins that do not have any known functionality or anatomical domains. They may contain coiled-coil features or intracellular areas that make it difficult to ascribe their function. Moreover, studying the purpose of molecules with unknown roles provides several benefits, including the capability to identify novel helical configurations of 3-D constructions, which enables the evaluation of distinctive categories and designs, along with the discovery of novel amino acid processes and also sequences. Such emerging domains can serve as potential targets for medicine in the next generations. Furthermore, phylogenetic profiling of proteins across different genomes can be utilized to predict their functions (Basu et al., 2011), and utilizing advanced techniques, such as mass spectrometry for identifying protein complexes and microarray analysis for profiling gene expression (Brown et al., 2000), and also systematic synthetic lethal analysis (Goehring et al., 2003) are useful. The concept behind clustering gene-expression patterns is that genes with related functions are more likely to be co-expressed (Yuan et al., 2008). To identify the function, researchers used the neighbor-counting method (Hu et al., 2010). They determined the function of an uncharacterized protein by analyzing the occurrence of neighboring proteins with particular activity. Instead of searching for a straightforward match between the roles of the interacting members, and a Bayesian technique was used to assess the probability of a hypothetical protein exhibiting the annotated function (Deng et al., 2002). A large number of protein domains have functions that are still unknown; nevertheless, these domains are involved in the chemical reactions of living things, which may lead to harmful effects. In certain instances, mutations, including additions, eliminations, and replacements, can alter the role of a protein. For instance, identification and functional annotation of hypothetical proteins in *Providencia rettgeri* strain MRSN845308 (Pal et al., 2024) *Vibrio parahaemolyticus* strain VP 128 (Mou et al., 2021), *Neisseria gonorrhoeae* (Lakhanpal et al., 2024), *and Neisseria meningitidis* (Asha et al., 2024) toward designing therapeutic targets.

Aeromonas veronii is a Gram-negative, rod-like, facultative anaerobic microorganism that can lead to infections in people. This organism is frequently found in various freshwater fish, surface water, and domestic animals. The most common diseases caused by *Aeromonas veronii* in humans include gastroenteritis, soft-tissue infections, and bacteremia (Liu et al., 2022). The primary goal of the research is to determine and characterize the E2102 protein domain from *Aeromonas veronii*, which has an unidentified function utilizing bioinformatics techniques.

Materials and Methods

Screening of Hypothetical Proteins (HP)

Hypothetical proteins were identified by searching term ‘hypothetical protein’ in the National Center for Biotechnology Information (NCBI) protein database, and matches were selected randomly to investigate their homologs applying BLAST tools. To determine the potential purposes of the hypothetical protein, a matching query was conducted utilizing NCBI blast software to detect genes with possible physical or functional homology (https://www.ncbi.nlm.nih.gov).

Physicochemical Characterization of the Hypothetical Proteins

The physicochemical properties of the raw hypothetical protein sequence were evaluated using the ProtParam tool available through the ExPASy service (Rahman et al., 2022). The software calculates and also delivers various metrics, including the molecular weight, theoretical isoelectric point (pI), amino acid composition, counts of positively and negatively charged residues, extinction coefficient, instability index, aliphatic index, and the grand average of hydropathicity (GRAVY), among others. Molar absorptivity of a molecule reflects the amount of electromagnetic radiation it receives at a certain frequency. The instability index provides an estimate of the protein's conformational integrity under laboratory conditions. An instability index below 40 suggests a stable, while a number more than 40 expresses instability. Aliphatic index represents the proportion of the area filled by aliphatic side of molecules, which contributes to protein thermostability. The GRAVY score is determined by dividing the sum of the hydropathy values of all amino acids by the number of residues in the sequence (https://web.expasy.org/protparam/).

Sequence Similarity

The fundamental step in predicting a protein's function involves identifying its sequence homologs within various genomic and proteomic databases. For this purpose, the widely used bioinformatics tool BLASTp was employed (Mahram & Herbordt, 2010).

Domain Identification and Gene Ontology Prediction

Firstly, various widely accessible computational resources and repositories, including Pfam, InterPro, CATH, SUPERFAMILY, SMART, SCANPROSITE, and CDD-BLAST, were utilized. Such computational platforms and biological data facilitate to identification of preserved

areas, enabling the classification of proteins based on their structural and functional properties. The functional roles of the hypothetical protein were interpreted based on similarity using Pfam (Mistry et al., 2021), InterPro (Blum et al., 2021), SUPERFAMILY (Gough et al., 2001), and SCANPROSITE (de Castro et al., 2006). Additionally, SMART and CATH were employed to identify the functions of our hypothetical proteins by analyzing their domain architecture and categorizing the domains within the structural hierarchy (Letunic et al., 2021; Sillitoe et al., 2015). The Conserved Domain Database (CDD) was used to identify conserved domains (Lu et al., 2020). Moreover, the gene ontology resource offers a structured and computable framework for understanding the functions of genes and their associated products. In this study, we also predicted the gene ontology of the hypothetical protein using Argot$^{2.5}$(Annotation Retrieval of Gene Ontology Terms) (Lavezzo et al., 2016).

Assessment of Secondary Structure

The secondary structural elements of the hypothetical protein (accession no. PXV57966.1) were predicted using the SOPMA tool (Combet et al., 2000) with default parameters, including a window width of 17, four states, and a similarity threshold of 8.

Determination of the Sub-cellular Localization

Determination of the intracellular localization of an amino acid provides crucial insights into its potential function. Proteins can localize to the outer membrane, inner membrane, periplasm, extracellular space, and cytoplasm depending on their functions and structure (Gazi et al., 2016). In this study, PSORTb (Yu et al., 2010) and CELLO (Sanchez, 2013) were used to predict the subcellular localization of the proteins. Furthermore, SOSUI (Hirokawa et al., 1998), HMMTOP (Tusnády & Simon, 2001), TMHMM (Krogh et al., 2001), and SignalP (Nielsen et al., 2019) were employed to predict transmembrane helices and identify signal peptide cleavage sites.

Protein-Protein Interaction Analysis

The role of proteins is determined by the interactions among their amino acid residues. These interactions influence the protein's structure, stability, and function, enabling it to carry out specific biological processes. In our study, we utilized the STRING database (http://string-db.org/), which integrates physical and functional associations to analyze and predict known and potential protein interactions (Szklarczyk et al., 2021). This relied on genetic environment, excellent efficiency of analyses, combination, and existing information. This database combines interaction data from the following sources in a quantitative manner (Franceschini et al., 2013).

Homology-based Protein Modeling Using HHPred

Traditional sequence query techniques explore sequence resources, like UniProt or unique datasets. HHpred (Söding et al., 2006) explores multiple databases, such as PDB, SCOP, Pfam, SMART, COGs, and CDD, to facilitate homology detection and protein structure prediction. HHpred is a high-performance platform that employs hidden Markov models (HMMs) pairing features to identify, predict distant peptide similarity and shape.

Results and Discussion

The Retrieval of Protein Sequences

The protein E2102 from *Aeromonas veronii*, whose function is unknown, was retrieved from the National Center for Biotechnology Information (NCBI) with the accession number PXV57966.1. The 3D structure of this protein is not contained in the Protein Data Bank (PDB). Therefore, a BLASTp analysis was conducted using the NCBI database to obtain initial information on the secondary and tertiary structures of the 363-amino-acid protein PXV57966.1 identified in *Aeromonas veronii*. The sequence of proteins PXV57966.1 was identified from the result using various computational approaches and tools to determine its prospective properties in use. The FASTA format of this sequence is shown in Figure 1.

Physicochemical Characterization

Comprehensive physicochemical analysis helps characterize a protein's structural and functional properties, including potential outer regions involved in biological activity. The amino acid sequence of *Aeromonas veronii* (accession no. PXV57966.1) was obtained in FASTA format and used as the sequence of queries for determining the physical and chemical parameters. PXV57966.1 has an instability index of 27.61, suggesting that the protein is stable (Guruprasad et al., 1990). The protein exhibits a molecular weight of 39.47 kDa and an isoelectric point (pI) of 4.81, indicating it is acidic. The GRAVY is a numerical scale used to estimate the overall hydrophobicity or hydrophilicity based on its amino acid composition (Jaspard et al., 2012). PXV57966.1 exhibited negative values, suggesting that it might be hydrophilic. The physical and chemical profile of the PXV57966.1 protein is provided in Table 1.

Sequence Analogy Assessment

The E2102 domain sequence (accession number PXV57966.1) was analyzed for sequence similarity using the BLASTp structural protein data file against a unique dataset. The percentage identity, common features, and corresponding E-values are presented in Table 2.

Analysis of Secondary Structures

The structure and function of proteins are closely associated. The secondary structure elements, including helix, coil, sheet, and turn, are crucial in determining a protein's structure, function, and interactions. The SOPMA tool predicted the protein's secondary structure to consist of 178 residues (49.04%) in the alpha helix (Hh), 59 residues (16.25%) in the extended strand (Ee), and 126 residues (34.71%) in the random coil (Cc) (Figure 2). The result showed the absence of other secondary structures such as 3_{10} helix, Pi helix, Beta bridge, Beta turn, Bend region, and ambiguous conditions (Figure 2).

Subcellular Localization

Determining a protein's subcellular location is crucial for identifying potential therapeutic and vaccination targets (Acharya & Garg, 2016). Proteins located in the cytoplasmic matrix could serve as potential drug targets, whereas proteins in the inner and outer membranes could be used as vaccination targets (Prabhu et al., 2020). Analyzing the distribution of proteins is fundamental to uncovering their functions. Based on the knowledge of trained data sets, the hypothetical protein was expected to be found in various cellular locations. The protein was anticipated to be found in the cytoplasm. Due to its localization in the

Protein Identifier

>PXV57966.1, hypothetical protein DFO51_101539 [*Aeromonas veronii*]

Amino Acid number: 363

Protein sequence

MGFKCGIVGLPNVGKSTLFNALTKAGIEAANFPFCTIEPNTGVVPMPDPRLDQLAAIINPQRVVPTTMEFVDIA
GLVAGASKGEGLGNQFLANIRETEAIGHVVRCFDDENIIHVAGKVSPADDIEVINTELALSDLDACERAIHRQSK
RAKGGDKDAKLEVETLEKIKVALENGQMIRGMKLDKEELAAVSHLNFLTLKPTMYIANVAEDGFENNPYLDKV
REIAAAENAVVVVVCCAIEADIAELDDEDRAEFMADLGIEEPGLNRVIRSGYQLLNLQTYFTAGVKEVRAWTIP
VGATAPQAAGKIHTDFEKGFIRAQTIAFEDFINYKGEQGAKEAGKMRAEGKDYIVKDGDIMNFLFNV

Figure 1. Hypothetical protein amino acid sequence in FASTA format.

Table 1. Physical and chemical profile of PXV57966.1 protein

Property	Value
Number of amino acids	363
Molecular weight	39473.11
Theoretical PI	4.81
Total number of negatively charged residues	56
Total number of positively charged residues	38
Ext. Coefficient	14815
Instability index	27.61
Aliphatic index	94.10
Grand average of hydropathicity (GRAVY)	-0.075

Table 2. The BLASTp screening results reveal homologous E2102 sequences in comparison with various distinct counterparts

Protein ID	Protein	Organism	Identity	Similarity	Score	E-value
WP_244776409.1	redox-regulated ATPase YchF	*Aeromonas veronii*	99.72%	100%	739	0.0
MFM5578876.1	redox-regulated ATPase YchF	*Aeromonas veronii*	99.45%	100%	739	0.0
MFM5433166.1	redox-regulated ATPase YchF	*Aeromonas veronii*	99.17%	100%	736	0.0
WP_033137132.1	redox-regulated ATPase YchF	*Aeromonas finlandensis*	99.17%	100%	735	0.0
WP_219271086.1	redox-regulated ATPase YchF	*Aeromonas jandaei*	99.45%	100%	735	0.0
WP_040098811.1	redox-regulated ATPase YchF	*Aeromonas australiensis*	99.17%	100%	735	0.0
WP_411662536.1	redox-regulated ATPase YchF	*Aeromonas enteropelogenes*	99.17%	100%	734	0.0
WP_411625073.1	redox-regulated ATPase YchF	*Aeromonas allosaccharophila*	98.90%	100%	734	0.0
WP_336288886.1	redox-regulated ATPase YchF	*Aeromonas dhakensis*	98.07%	100%	727	0.0
WP_042011576.1	redox-regulated ATPase YchF	*Aeromonas fluvialis*	97.80%	100%	727	0.0

cytoplasm, no results were obtained from TMHMM, HMMTOP, and SignalP servers. Predicting membrane proteins is crucial for understanding potential drug targets and designing effective therapeutic compounds (Mou et al., 2021).

Domain Identification and Gene Ontology Prediction

We determined the domain of the Hypothetical protein, that represents fundamental, useful, and genetic components of a molecule, thereby offering insight into its functional role. In our study, the protein sequence PXV57966.1 belongs to the Ribosome-binding ATPase YchF. Moreover, the Ribosome-binding ATPase YchF domain is essential for ribosome biogenesis and translational control, ensuring proper ribosome assembly and regulating protein synthesis. Its ATPase activity aids ribosomal function, supporting cellular homeostasis and stress adaptation by modulating translation (Landwehr et al., 2022). Furthermore, Gene Ontology (GO) plays a crucial role in understanding protein function by categorizing genes and their products based on biological processes, molecular functions, and cellular components. Gene ontology was used to examine and categorize the functional aspects of the biological process ontology (Table 3), molecular function ontology (Table 4), and cellular component ontology (Table 5).

```
SOPMA :
   Alpha helix      (Hh) :   178 is  49.04%
   3_10  helix      (Gg) :     0 is   0.00%
   Pi helix         (Ii) :     0 is   0.00%
   Beta bridge      (Bb) :     0 is   0.00%
   Extended strand  (Ee) :    59 is  16.25%
   Beta turn        (Tt) :     0 is   0.00%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :   126 is  34.71%
   Ambiguous states (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```

Figure 2. SOPMA analysis results for secondary structure elements.

Table 3. Biological process ontology

GO ID	GO Term
GO:0055114	oxidation-reduction process
GO:0090502	RNA phosphodiester bond hydrolysis, endonucleolytic
GO:0006414	translational elongation

Table 4. Molecular function ontology

GO ID	GO Term
GO:0043023	ribosomal large subunit binding
GO:0016887	ATPase activity
GO:0043022	ribosome binding
GO:0005525	GTP binding
GO:0000166	nucleotide binding
GO:0005524	ATP binding
GO:0016787	hydrolase activity
GO:0016491	oxidoreductase activity
GO:0003676	nucleotide binding

Table 5. Component of the cell ontology

GO ID	GO Term
GO:0005737	cytoplasm
GO:0044444	cytoplasmic part

Study of Protein-Protein Relationships

In our study, PXV57966.1 interacts with peptidyl-tRNA hydrolase with a high confidence score of 0.957. Moreover, peptidyl-tRNA hydrolase is essential for maintaining efficient and accurate protein synthesis. It resolves stalled ribosomes by removing incomplete polypeptides attached to tRNA, preventing toxic accumulation and ensuring the proper functioning of the translation machinery. The protein-protein interaction of the hypothetical protein is illustrated in Figure 3.

Determination of Homology

To create a remote homolog model, HHpred was utilized. The HHpred result (Figure 4), new similar sequence to the E2102 protein was identified, including GTP-binding protein, YchF protein, GTPase family protein, YchF GTP-binding protein, Ferrous iron transport protein B, hydrogenase expression/formation protein. A homology-derived model was generated from HHpred at the Modeler server using 7Y9I (a known structure from PDB) as the template. The reliability of the model was evaluated using the VERIFY3D and ProSA-web server tools. The overall quality factor determined by ERRAT was 97.27%. Additionally, two reference lines are displayed on the error graph to indicate the confidence thresholds beyond which regions can be considered unreliable. The percentage represents the portion of the protein with an estimated error value below the 95% rejection limit. High-resolution protein structures generally show values around 95% or higher, whereas those with lower resolution (between 2.5 and 3 Å) usually yield values close to 91% (Figure 5). Furthermore, the ProSA-web server is employed to evaluate the confirmation efficiency and identify possible errors in a fundamental tertiary structure model. The final E2102 protein model was validated with a Z-score of -8.04, indicating that the model is significant (Figure 6). The 7Y9I template model was visualized using ChimeraX 1.9 (Figure 7).

Conclusion

The analysis indicated that the Ribosome-binding ATPase YchF domain is an essential to pharmaceutical goal due to its ability to bind to ribosomal subunits. It was also discovered that E2102 is a soluble protein with a single exposed domain. The presence and distribution of E2102 domains across a wide variety of bacteria and diseases suggest the potential for developing new antibacterial drugs. Ongoing studies aim to identify the active site region of the modeled protein, with protein-ligand docking analyses being carried out to determine the specific amino acid residues involved in ligand binding.

Ethical Statement

The authors confirm that no animal experiments were carried out as part of this research.

Funding Information

No external funding was provided for this study.

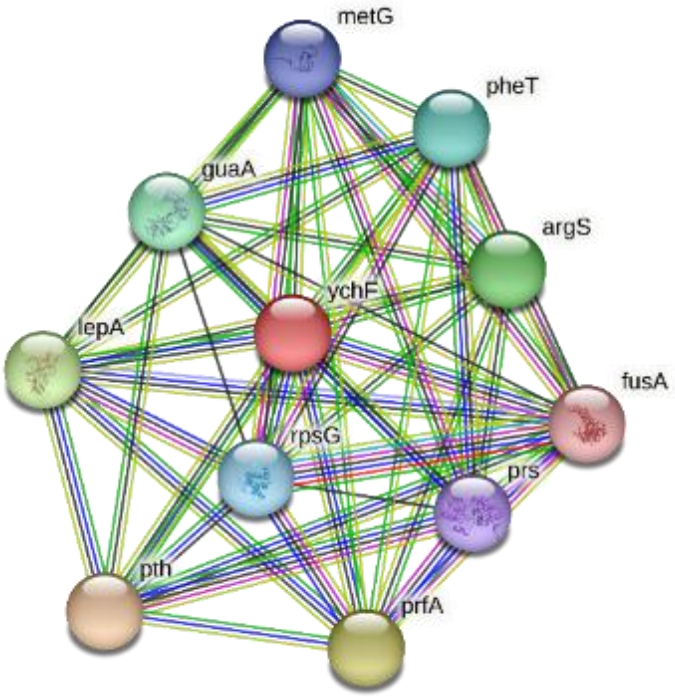

Figure 3. Protein interaction analysis of hypothetical analysis using the String database, represented as ychF.

```
Query         PXV57966.1, hypothetical protein DFO51_101539 [Aeromonas veronii]
Match_columns 363
No_of_seqs    1 out of 3
Neff          1
Searched_HMMs 70004
```

No	Hit	Prob	E-value	P-value	Score	SS	Cols	Query HMM	Template HMM
1	7Y9I_A Obg-like ATPase 1; AtYc	100.0	2.2E-70	3.1E-75	460.4	17.7	361	2-363	24-388 (394)
2	1NI3_A YchF GTP-binding protei	100.0	2E-69	2.9E-74	451.1	18.2	361	2-363	20-391 (392)
3	2OHF_A GTP-binding protein 9;	100.0	2.5E-66	3.6E-71	434.6	18.5	361	2-363	22-389 (396)
4	1JAL_B YchF protein; nucleotid	100.0	1.9E-61	2.7E-66	392.4	20.5	363	1-363	1-363 (363)
5	2DBY_A GTP-binding protein; GT	100.0	7.2E-60	1E-64	390.8	20.1	360	3-363	2-368 (368)
6	1WXQ_A GTP-binding protein; GT	100.0	7E-47	9.9E-52	321.9	10.4	336	3-363	1-396 (397)
7	7NRC_So Ribosome-interacting G	100.0	2.8E-29	4E-34	204.2	6.0	282	3-359	63-364 (366)
8	4A9A_B RIBOSOME-INTERACTING GT	99.5	2.4E-14	3.4E-19	118.1	8.2	278	3-362	73-376 (376)
9	4CSU_9 GTPASE OBGE/CGTA; (P)PP	99.1	8.6E-11	1.2E-15	99.2	4.1	88	3-106	160-247 (390)
10	5M04_A GTPase ObgE/CgtA; GTPas	99.0	7.3E-10	1E-14	93.7	4.2	90	3-108	180-269 (360)
11	1UDX_A the GTP-binding protein	98.9	1.6E-09	2.3E-14	92.4	4.6	90	3-108	158-247 (416)
12	7OI6_y GTP-binding protein 10;	98.9	2.2E-09	3.2E-14	92.3	4.8	88	3-106	149-236 (387)
13	1LNZ_A SPO0B-associated GTP-bi	98.9	2.9E-09	4.2E-14	88.5	4.9	93	3-111	159-251 (342)
14	7OF7_x Mitochondrial ribosome-	98.8	9.7E-09	1.4E-13	90.7	4.5	92	3-110	225-316 (406)
15	9BA6_B Ferrous iron transport	98.7	3.3E-08	4.7E-13	73.4	3.9	92	3-111	3-96 (261)
16	8EWH_B 50S ribosomal subunit a	98.6	2.3E-09	3.3E-14	99.5	-3.8	244	3-270	6-288 (607)
17	3B1V_A Ferrous iron uptake tra	98.2	1.1E-06	1.6E-11	66.3	2.3	45	1-45	2-46 (272)
18	7YLA_6 GTPase HflX; HflX, RIBO	98.1	9.7E-06	1.4E-10	72.0	5.5	89	2-108	197-285 (426)
19	8KAB_h GTPase HflX; Complex, R	98.0	7.5E-06	1.1E-10	75.0	4.0	91	3-111	246-336 (483)
20	3W5J_A Ferrous iron transport	98.0	1.5E-05	2.1E-10	56.2	4.4	87	3-108	4-90 (204)
21	3K53_B Ferrous iron transport	98.0	4.6E-06	6.5E-11	64.4	1.7	51	1-51	2-52 (271)
22	3A1S_B Iron(II) transport prot	97.9	5.6E-06	8E-11	67.9	1.5	49	3-51	6-54 (258)
23	3IBY_D Ferrous iron transport	97.9	1.4E-05	1.9E-10	58.1	2.4	46	3-48	2-47 (256)
24	3A1W_A Iron(II) transport prot	97.8	3.8E-05	5.5E-10	49.4	4.0	90	3-111	6-95 (168)
25	2WSM_B HYDROGENASE EXPRESSION/	97.8	4.3E-05	6.1E-10	53.5	3.6	38	3-42	31-68 (221)
26	8Q6Q_D Immunity-related GTPase	97.8	2.8E-05	4.1E-10	58.1	2.8	49	3-51	8-56 (192)
27	3H2Y_A GTPase family protein;	97.7	5.2E-05	7.4E-10	66.0	4.0	96	2-117	160-262 (368)

Figure 4. HHpred output displaying probable similarities with PXV57966.1.

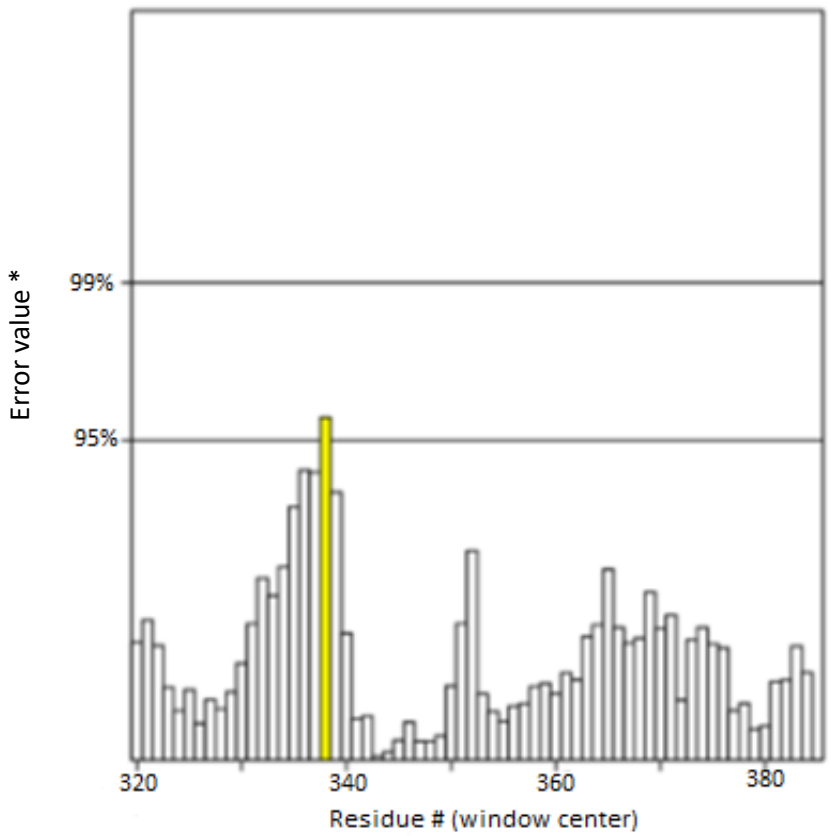

Figure 5. ERRAT evaluates the accuracy of the model.

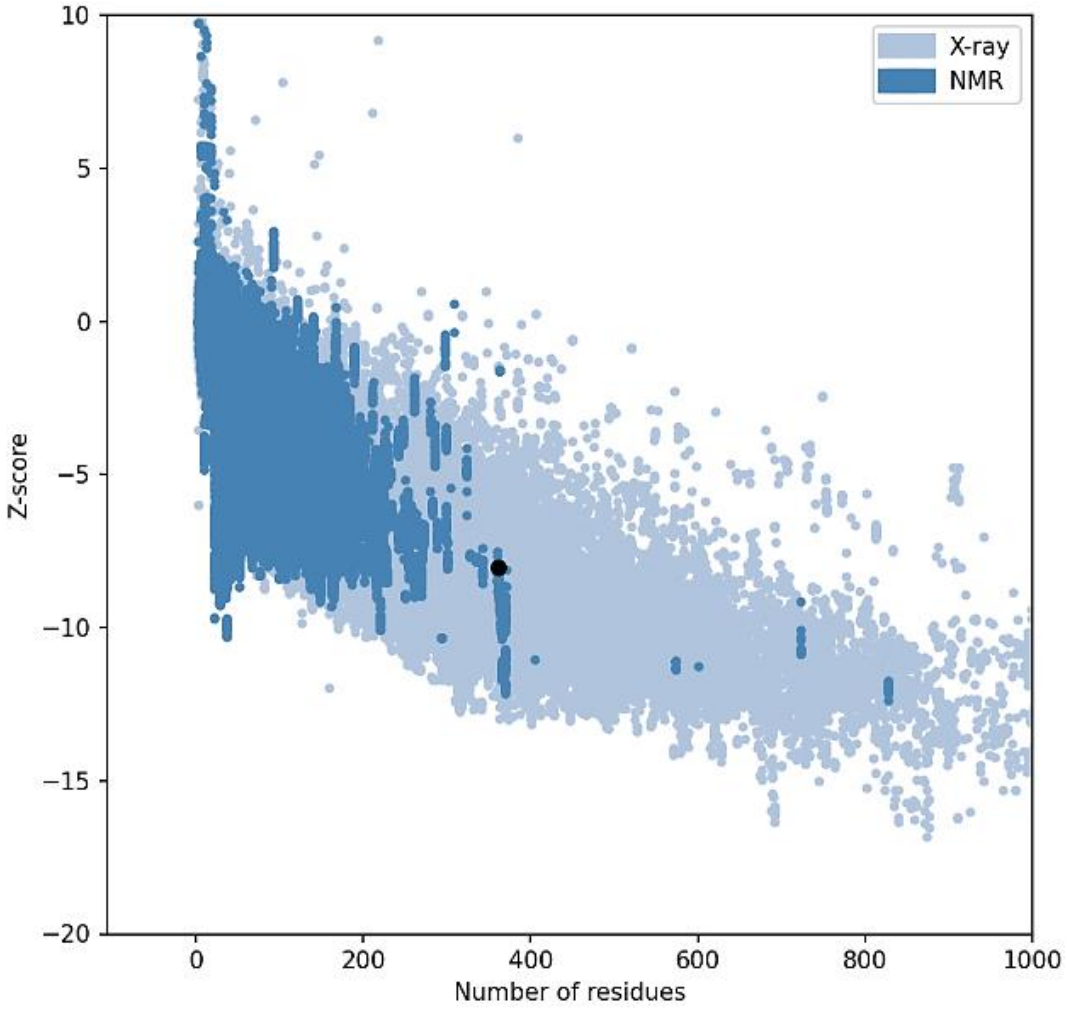

Figure 6. The model quality obtained by the ProSA-web server.

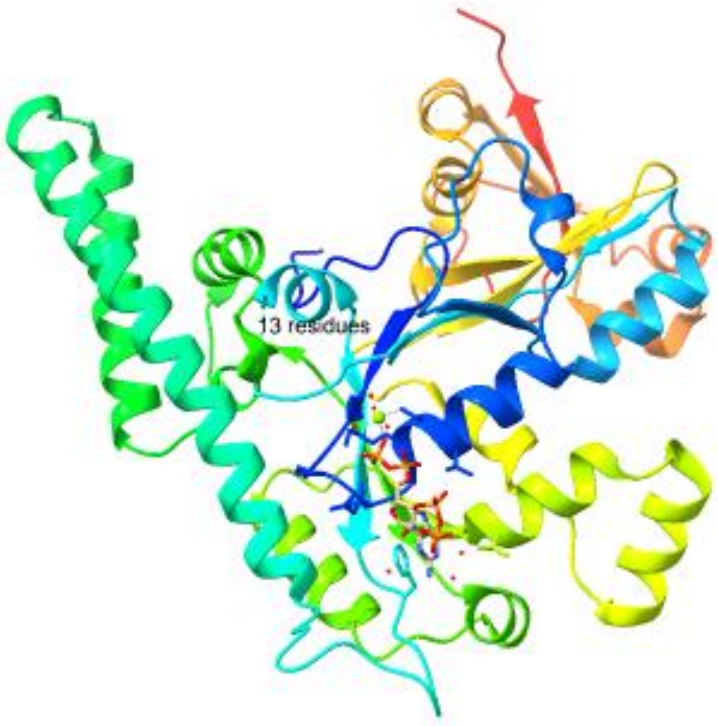

Figure 7. Visualization of the 7Y9I template model.

Author Contribution

Conceptualization, M.I.A., and S.I.M.; methodology, M.I.A., and S.I.M.; software M.I.A.; validation, M.I.A., and S.I.M.; formal analysis, M.I.A., and M.M.I.E.; investigation, M.I.A.; resources, M.I.A.; data curation, M.I.A., and S.I.M.; writing— original draft preparation, M.I.A.; writing-review and editing, M.I.A. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest

The author(s) declare that they have no conflicts of interest, whether financial, personal, or professional, that could be perceived as influencing the research in this paper.

Acknowledgements

The author thanks Sk Injamamul Islam and Sarower Mahfuj for helpful comments.

References

Acharya, A., & Garg, L. C. (2016). Drug Target Identification and Prioritization for Treatment of Ovine Foot Rot: An In Silico Approach. *International Journal of Genomics*, *2016*. https://doi.org/10.1155/2016/7361361

Asha, I. J., Gupta, S. D., Hossain, M. M., Islam, M. N., Akter, N. N., Islam, M. M., & Barman, D. N. (2024). In silico Characterization of a Hypothetical Protein (PBJ89160. 1) from Neisseria meningitidis Exhibits a New Insight on Nutritional Virulence and Molecular Docking to Uncover a Therapeutic Target. *Evolutionary Bioinformatics*, *20*, 11769343241298307. http://doi.org/1177/11769343241298307

Basu, M. K., Selengut, J. D., & Haft, D. H. (2011). *ProPhylo partial phylogenetic profiling to guide protein family construction and assignment of biological process*. BMC bioinformatics, 12, 434-434. https://doi.org/10.1186/1471-2105-12-434

Bharat Siva Varma, P., Adimulam, Y. B., & Kodukula, S. (2015). In silico functional annotation of a hypothetical protein from Staphylococcus aureus. *Journal of Infection and Public Health*, *8*(6), 526–532. https://doi.org/10.1016/j.jiph.2015.03.007

Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., Finn, R. D. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, *49*(D1), D344–D354. https://doi.org/10.1093/nar/gkaa977

Brown, M. P. S., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., Ares, M., & Haussler, D. (2000). *Knowledge-based analysis of microarray gene expression data by using support vector machines*. *Proceedings of the National Academy of Sciences, 97(1), 262-267*. https://doi.org/10.1073/pnas.97.1.262

Combet, C., Blanchet, C., Geourjon, C., & Deléage, G. (2000). NPS@: Network protein sequence analysis. *Trends in Biochemical Sciences*, *25*(3), 147–150. https://doi.org/10.1016/S0968-0004(99)01540-6

de Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A., & Hulo, N. (2006). ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research*, *34*(WEB. SERV. ISS.), 362–365. https://doi.org/10.1093/nar/gkl124

Deng, M., Zhang, K., Mehta, S., Chen, T., & Sun, F. (2002, August). Prediction of protein function using protein-protein interaction data. Proc IEEE Comput Soc Bioinform Conf, 1, 197-206. https://doi.org/10.1089/106652703322756168

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., Von Mering, C., & Jensen, L. J. (2013). STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research*, *41*(D1), 808–815. https://doi.org/10.1093/nar/gks1094

Galperin, M. Y., & Koonin, E. V. (2004). *Conserved hypothetical proteins prioritization of targets for experimental study*. *Nucleic acids research*, 32(18), 5452-5463. https://doi.org/10.1093/nar/gkh885

Gazi, M. A., Kibria, M. G., Mahfuz, M., Islam, M. R., Ghosh, P., Afsar, M. N. A., Khan, M. A., & Ahmed, T. (2016). Functional, structural and epitopic prediction of hypothetical proteins of Mycobacterium tuberculosis H37Rv: An in silico approach for prioritizing the targets. *Gene*, *591*(2), 442–455. https://doi.org/10.1016/j.gene.2016.06.057s

Goehring, A. S., Mitchell, D. A., Hin, A., Tong, Y., Keniry, M. E., Boone, C., & Sprague, G. F. (2003). *Synthetic Lethal Analysis Implicates Ste20p, a p21- activated Protein Kinase, in Polarisome Activation*. *Molecular biology of the cell,* 14(4), 1501-1516. https://doi.org/10.1091/mbc.e02-06-0348

Gough, J., Karplus, K., Hughey, R., & Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of Molecular Biology*, *313*(4), 903–919. https://doi.org/10.1006/jmbi.2001.5080

Guruprasad, K., Reddy, B. V. B., & Pandit, M. W. (1990). Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering, Design and Selection*, *4*(2), 155–161. https://doi.org/10.1093/protein/4.2.155

Hirokawa, T., Boon-Chieng, S., & Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, *14*(4), 378–379. https://doi.org/10.1093/bioinformatics/14.4.378

Hu, P., Jiang, H., & Emili, A. (2010). Predicting protein functions by relaxation labelling protein interaction network. *BMC bioinformatics*, 11, 1-8. https://doi.org/10.1186/1471-2105-11-S1-S64

Jaspard, E., Macherel, D., & Hunault, G. (2012). Computational and statistical analyses of amino acid usage and physico-chemical properties of the twelve Late Embryogenesis Abundant Protein classes. Plos One, *7*(5), e36968. https://doi.org/10.1371/journal.pone.0036968

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete

genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Lakhanpal, G., Tiwari, H., Shukla, M. K., & Kumar, D. (2024). In silico exploration of hypothetical proteins in Neisseria gonorrhoeae for identification of therapeutic targets. *In Silico Pharmacology*, *12*(1), 10. https://doi.org/10.1007/s40203-023-00186-w

Landwehr, V., Milanov, M., Hong, J., & Koch, H. G. (2022). The role of the universally conserved atpase ychf/ola1 in translation regulation during cellular stress. *Microorganisms*, *10*(1), 1–16. https://doi.org/10.3390/microorganisms10010014

Lavezzo, E., Falda, M., Fontana, P., Bianco, L., & Toppo, S. (2016). Enhancing protein function prediction with taxonomic constraints The Argot2.5 web server. *Methods*, *93*(2016), 15–23. https://doi.org/10.1016/j.ymeth.2015.08.021

Letunic, I., Khedkar, S., & Bork, P. (2021). SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Research*, *49*(D1), D458–D460. https://doi.org/10.1093/nar/gkaa937

Liu, F., Yuwono, C., Chin, A., Tay, Y., Wehrhahn, M. C., & Riordan, S. M. (2022). Analysis of global Aeromonas veronii genomes provides novel information on source of infection and virulence in human gastrointestinal diseases. *BMC Genomics*, 1–15. https://doi.org/10.1186/s12864-022-08402-1

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Research*, *48*(D1), D265–D268. https://doi.org/10.1093/nar/gkz991

Lubec, G., Afjehi-Sadat, L., Yang, J. W., & John, J. P. P. (2005). Searching for hypothetical proteins: Theory and practice based upon original data and literature. *Progress in Neurobiology*, *77*(1–2), 90–127. https://doi.org/10.1016/j.pneurobio.2005.10.001

Mahram, A., & Herbordt, M. C. (2010). Fast and accurate NCBI BLASTP: acceleration with multiphase FPGA-based prefiltering. Mahram, A., & Herbordt, M. C. (2010, June). Fast and accurate NCBI BLASTP: Acceleration with multiphase FPGA-based prefiltering. In Proceedings of the 24th ACM International Conference on Supercomputing (pp. 73-82). https://doi.org/10.1145/1810085.1810099

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913

Mou, M., Islam, S., & Mahfuj, M. (2021). In Silico Functional Annotation of VP 128 Hypothetical Protein from Vibrio parahaemolyticus. *Aquatic Food Studies*, *1*(2), 1-12. https://doi.org/10.4194/AFS37

Nielsen, H., Tsirigos, K. D., Brunak, S., & von Heijne, G. (2019). A Brief History of Protein Sorting Prediction. *Protein Journal*, *38*(3), 200–216. https://doi.org/10.1007/s10930-019-09838-3

Pal, D. C., Anik, T. A., Rahman, A. A., & Mahfujur Rahman, S. M. (2024). Identification and Functional Annotation of Hypothetical Proteins of Pan-Drug-Resistant Providencia rettgeri Strain MRSN845308 Toward Designing Antimicrobial Drug Targets. *Bioinformatics and Biology Insights*, *18*, 11779322241280580. https://doi.org/10.1177/11779322241280580

Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics*, *42(1), 3-1*. https://doi.org/10.1002/0471250953.bi0301s42

Prabhu, D., Rajamanikandan, S., Anusha, S. B., Chowdary, M. S., Veerapandiyan, M., & Jeyakanthan, J. (2020). In silico Functional Annotation and Characterization of Hypothetical Proteins from Serratia marcescens FGI94. *Biology Bulletin*, *47*(4), 319–331. https://doi.org/10.1134/S1062359020300019

Rahman, M. A., Heme, U. H., & Parvez, M. A. K. (2022). In silico functional annotation of hypothetical proteins from the *Bacillus paralicheniformis* strain Bac84 reveals proteins with biotechnological potentials and adaptational functions to extreme environments. Plos One, *17*(10 October), 1–22. https://doi.org/10.1371/journal.pone.0276085

Reichart, N. J., Jay, Z. J., Hatzenpichler, R., Krukenberg, V., Parker, A. E., & Spietz, R. L. (2020). Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate amendment. *The ISME Journal*, 2851–2861. https://doi.org/10.1038/s41396-020-00749-1

Sanchez, G. (2013). Las instituciones de ciencia y tecnología en los procesos de aprendizaje de la producción agroalimentaria en Argentina. *El Sistema Argentino de Innovación: Instituciones, Empresas y Redes. El Desafío de La Creación y Apropiación de Conocimiento.*, *14*(October 2006), 659–664. https://doi.org/10.1002/prot

Sillitoe, I., Lewis, T. E., Cuff, A., Das, S., Ashford, P., Dawson, N. L., Furnham, N., Laskowski, R. A., Lee, D., Lees, J. G., Lehtinen, S., Studer, R. A., Thornton, J., & Orengo, C. A. (2015). CATH: Comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Research*, *43*(D1), D376–D381. https://doi.org/10.1093/nar/gku947

Söding, J., Remmert, M., Biegert, A., & Lupas, A. N. (2006). HHsenser: Exhaustive transitive profile search using HMM-HMM comparison. *Nucleic Acids Research*, *34*(WEB. SERV. ISS.), 374–378. https://doi.org/10.1093/nar/gkl195

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., & von Mering, C. (2021). The string database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research*, *49*(D1), D605-D612. https://doi.org/10.1093/nar/gkaa1074

Tusnády, G. E., & Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics*, *17*(9), 849–850. https://doi.org/10.1093/bioinformatics/17.9.849

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Cenk Sahinalp, S., Ester, M., Foster, L. J., & Brinkman, F. S. L. (2010). PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, *26*(13), 1608–1615. https://doi.org/10.1093/bioinformatics/btq24