INNOVATIONS IN MOLECULAR BIOTECHNOLOGY
Jul-Dec 2025, VOL. 3, ISSUE 2, pp. 1-5
https://doi.org/10.61577/imb.2025.100006

RESEAPRO JOURNALS

ORIGINAL ARTICLE

OPEN ACCESS

# Computational Identification of Hypothetical Proteins in the Core Proteome of *Orientia tsutsugamushi*

**Subrat Kumar Swain**

**Medical Research Laboratory, IMS and SUM Hospital, Siksha 'O' Anusandhan deemed to be University, Bhubaneswar, Odisha, India**

## ABSTRACT

Advanced technologies have generated a wealth of microbial genomics data, resulting in numerous uncharacterized genes. A large portion of the *Orientia tsutsugamushi* proteome consists of such hypothetical and conserved proteins with poorly understood functions. In this study, we chose 292 hypothetical proteins from *O. tsutsugamushi* strains (Ikeda and Boryong) for systematic in-silico characterization. We used a series of computational tools (ExPASy ProtParam, VirulentPred, PSORTb, Pfam, InterProScan, PSIPRED, SWISS-MODEL, CASTp, and STRING) to predict physicochemical properties, subcellular localization, functional domains, secondary and tertiary structure, active sites, and protein–protein interactions. Comparative proteomic analysis identified stable proteins, cytoplasmic proteins, and virulent proteins. Notably, proteins that showed no significant homology to the human proteome but are likely key to bacterial function. Functional classification placed many hypothetical proteins at the enzyme class (e.g., phosphatases, transferases, hydrolases, kinases, peptidases, isomerases, ligases, oxidoreductases), transporters, binding proteins, secretory proteins, and regulatory factors. We generated homology models and validated models with multiple structural assessment metrics and identified areas of activity of interest for possible drug targeting.Our in-silico annotations provide hypotheses for the function of previously uncharacterized proteins of *O. tsutsugamushi* and provide a ranked set of prioritized targets for experimental laboratory validation and structure-based drug design.

## KEY WORDS

*Orientia tsutsugamushi*; Hypothetical protein; In- silico analysis; Functional characterization

## ARTICLE HISTORY

Received 1 July 2025;
Revised 22 July 2025;
Accepted 28 July 2025

## Introduction

The acute, mite-borne zoonosis scrub typhus is caused by the obligate intracellular bacterium *Orientia tsutsugamushi* (*O. tsutsugamushi*), causing a spectrum of diseases in infected patients, from self-limited febrile illnesses to more serious and potentially fatal manifestations, such as pneumonitis, acute respiratory distress syndrome, myocarditis, renal impairment, and neurological involvement [1]. Although scrub typhus has historically been limited to the Asia–Pacific "tsutsugamushi triangle," recent independent sources of epidemiological data have shown that human cases are being reported outside of this area, and the significance and public health impact are being recognized in various regions [2]. The infections have been documented in specific geographical areas of India, and we have associated scrub typhus infection with a range of sequelae and mortality outcomes, including a death rate of 30% or more [2-5]. The predominant symptom is high and continuous fever. Other common recorded temperatures are greater than 101°F for more than 1 week [6]. Most incubation periods are variable from 6 to 21 days episodic mostly around 10 to 12 days. If not treated, infected patients have developed serious manifestations in the second week, including pneumonitis, ARDS, myocarditis, renal impairment, and neurologic involvement [7].

Genomic research of *O. tsutsugamushi* has shown that it has a unique and complex genome that is characterized by a high proportion of repetitive elements, frequent fragmentation of genes, and a significantly high number of predicted coding sequences currently annotated as hypothetical. Hypothetical proteins (HPs) comprise a significant fraction of predicted proteomes within other bacterial genomes, representing a substantial reservoir of uncharacterized biology. A systematic analysis of these proteins has the potential to identify new virulence factors, key components of metabolism, and pathogen-specific molecules that can be used as targets for treatment [8]. The Boryong strain contains 963 of 2,179 protein-coding sequences as fragmented genes, an astounding 49.6% of its predicted coding capacity [9]. The predicted proteomes for strains Ikeda and Boryong are 2,186 and 2,443 proteins, respectively, however the complete proteomic level characterization is poor. Understanding protein structure and function is crucial for elucidating the roles of virulence factors and to determining the mechanisms that enable microbial life [10]. Open reading frames that are conserved but have no experimentally validated function are usually referred to as "hypothetical proteins" (HPs). HPs are presumed to be encode proteins based on the overall sequence, but experimentally validating their biochemical characteristics and biological roles is required [11]. HPs can sometimes be regarded as "unknowns", since it is unclear whether they actually encode functional proteins; HPs are estimated to constitute roughly 30–40% of many bacterial genomes [12, 13]. Computational methods offer a fast and low-cost way to infer likely functions, cellular localizations, structural characteristics, and interaction partners of HPs to differentiate and rank best candidates for experimental work. In this paper, we presented a holistic in-silico pipeline to annotate specific hypothetical proteins of the core proteomes of *O. tsutsugamushi* (Ikeda and Boryong strains). By utilizing physicochemical profiling, subcellular localization, domain/family assignment, secondary and tertiary structure prediction, active site mapping, and host non-homology screening, we aimed to identify pathogenic specific proteins that may be involved in virulence or essential physiology and generate a rank list to guide biochemical studies and structure-based drug design.

## Materials and Methods

### Sequence retrieval

The proteome and genome sequences for *O. tsutsugamushi* strains are found on NCBI and UniProt. We collected FASTA sequences and accession identifiers for proteins from Ikeda and Boryong, giving datasets of 1,436 and 1,330 proteins, with 292

***Correspondence:** Mr. Subrat Kumar Swain, Medical Research Laboratory, IMS and SUM Hospital, Siksha 'O' Anusandhan deemed to be University, Bhubaneswar, Odisha, India, email: subratswain333@gmail.com

(Ikeda) and 333 (Boryong) annotations as hypothetical proteins respectively. Short sequences (<100 amino acids) and apparent pseudogenes were removed to limit false positives.

### Physicochemical properties and virulent factor prediction

We utilized ExPASy ProtParam to determine some physicochemical parameters of the molecule (molecular weight, extinction coefficient, aliphatic index, instability index, theoretical pI, GRAVY). We used SVM-based tools (VirulentPred, VICMpred) for predicting virulence potential [14].

### Protein localization and transmembrane region identification

Sub-cellular localization prediction was accomplished with PSORTb, CELLO2GO, and PSLpred; transmembrane helices and membrane vs soluble status with SOSUI. Signal peptides and cleavage sites were predicted with SignalP 4.1 [15]. PSLpred predicts subcellular localization for only Gram-negative bacteria and is therefore more specific. SOSUI will discriminate between the amino acid sequences of membrane proteins and soluble proteins and predict transmembrane helices [16]. Signal peptide and the location of a cleavage site in the peptide chain were predicted by SignalP 4.1 [17].

### Protein's functional domain/ prediction and evaluation

Protein domains and families were identified with the Conserved Domain Database (CDD), by using InterProScan, SMART and Pfam. Gene Ontology (GO) terms and anticipated molecular/biological functions were inferred using Argot2, as well as other similar annotation tools [18].

### Protein structure analysis and validation

The secondary structure predictions were performed using PSIPRED and SOPMA. In terms of homology inference, BLASTp was performed against the PDB to select several templates, then the model was built using SWISS-MODEL. Validation was performed using PROCHECK, VERIFY3D, ERRAT, and ProSA-web (Z-scores), and Ramachandran plots (RAMPAGE) were also performed during the validation. UCSF Chimera and PyMOL were used for visualisation. For the quality validation of homology models, various evaluation tools were implemented. The resulting HPs 3D structures were validated by PROCHECK [19], VERIFY 3D [20] and ERRAT tools. The Z value of the template and the target protein was determined using the ProSA-web server [21], and 3D structure graphics were made as movies using UCSF Chimera and PyMOL.

### Non-homology analysis

To identify pathogen-specific targets, predicted HP sequences with confident functional annotations were subjected to BLASTp against the human proteome (e-value cutoff 1e-4; bit-score threshold 100). Proteins lacking significant human homologues were prioritized [22].

## Result

### Sequence analysis

Hypothetical proteins (HPs) shorter than 100 amino acids were excluded from analysis. For the included strains Ikeda and Boryong, 36 and 27 HPs had lengths >100 aa, and met the other inclusion criteria for predicted stability and virulence potential. All retained hypothetical protein sequences were submitted to multiple functional-annotation servers for biological assignment; we excluded pseudogenes (i.e. proteins <100 aa) to minimize the risk of misannotation. After excluding HPs without supportive evidence, of the original predictions of 292 and 333 HPs, the HPs that met all selection criteria were 36 (Ikeda) and 27 (Boryong).

The physicochemical analysis by ProtParam showed considerable variability in molecular weights (~11 kDa to >262 kDa) and theoretical isoelectric points (pI 3.88–10.47). The extinction coefficients at 280 nm were significant, ranging from 1,490 to 142,605, which was due to the differences in Cys, Trp and Tyr content, but they provided useful values for quantitative studies of protein–protein and protein–ligand interaction studies. Instability indices for the chosen HPs paralleled one another at 10.33 to 39.99; the predicted stability of a protein is defined as being <40. GRAVY scores were comparatively moderate (~–1.098 to –0.845), but ~31% of proteins had positive GRAVY scores (hydrophobic). A low GRAVY score implies high hydrophilicity. Subcellular localization predictions indicated that the majority of the randomised HPs were classified as cytoplasmic proteins (Ikeda; ~62–63%, Boryong: ~52%), with a range of smaller predicted subsets identified as membrane-associated proteins and extracellular/released autotransporters.

The predicted localizations for the random HPs analysed from Ikeda were – cytoplasmic (22), cytoplasmic membrane (4), outer membrane (3), inner membrane (2), extracellular (3), and periplasmic (1); and for Boryong, predicted counts were-cytoplasmic (14), cytoplasmic membrane (3), outer membrane (6), inner membrane (1), and extracellular (3). SOSUI also identified a range of proteins with transmembrane helices in there predicted localizations, noting that there were up to two helices in an Ikeda protein (OTT_0001; sequence: "IFLVIIVAIINFITVANATASIC" and "LFKALWIFGIIIVLSVLAIKTTI"), while some Boryong proteins (OTBS_0978, OTBS_2105, OTBS_0302) had as many as three transmembrane helices.

### Functional annotation

Proteins usually have one or more functional and/or structural regions, known as domains. Conserved domains are a core functional unit in proteins, i.e., evolutionary Lego blocks that recombine in different ways in proteins to generate different functions. Of the 14 hypothetical proteins (HPs) examined, only one, OTT_0680 (a bacterial SH3 domain, a type of domain belonging to the SH3 domain superfamily and described under the Src homology-3 clan), contained specific domains. The remainder displayed unremarkable domains. 66 HPs were functionally annotated with high confidence across thirteen categories. The thorough understanding of these proteins is key to delineating the molecular mechanisms of pathogenesis and host-pathogen interactions. The HPs we identified have an enormous amount of enzymatic activity, such as phosphatases, transferases, hydrolases, kinases, peptidases, dehydrogenases, isomerases, nucleases, lyases, ligases, oxidases, and endonucleases. Others were classified as secretory, regulatory, metabolic, or transport.

The functional characterization of hypothetical proteins (HPs) demonstrated different enzymatic and/or regulatory functions that might contribute to bacterial survival and pathogenesis, given their respective roles for the bacterium. For example, phosphatases (OTT_0636, OTBS_1498) contributed to enhanced virulence by depleting phosphate in the host and/or organism; while transferases (11 HPs) have a role in acyl group transfer and glycosylation, which may be important for host-pathogen interaction. Four hydrolases (4 HPs) contributed to peptidoglycan cleavage and reformation, and four kinases (4 HPs) mediated signal transduction processes, the cell cycle, and post-translational modifications.



https://doi.org/10.61577/imb.2025.100006

Peptidases (5 HPs) had a role in the hydrolysis of peptides and possibly the processing and regulation of proteins. Isomerase (OTT_0464) supported cell wall biosynthesis via racemization of alanine versus cyclic ligases (OTT_0680, OTT_0309), which support the availability of RNA responsible for protein translation and metabolic process. Oxidoreductase proteins (OTT_1324) contributed to metabolic regulation and stress response; regulatory proteins (6 HPs) mediated transcriptional regulation, replication and proteolysis that may affect bacterial growth and pathogenesis. Three secretory proteins (3 HPs) were linked to exporting proteins and folding; thirteen binding proteins (13 HPs) were associated with binding DNA, ion, ATP, and NAD; and two transporter proteins (2 HPs) for the transportation of nutrients or ions. The functional characterization of HPs demonstrates that they have multi-functional roles in physiology and/or infection of bacteria.

## Structure analysis and assessment

In the Ikeda strain, the hypothetical proteins (HPs) OTT_0508, OTT_1526, and OTT_0464 were mostly arranged in extended α-helices. For the hypothetical proteins identified in the Boryong strain (OTBS_0057, OTBS_0096, OTBS_0161, OTBS_0806, OTBS_0856, and OTBS_1110), α-helical structures were present along with coiled-coil domains and β-sheets. Other proteins, like OTT_0690, OTT_0732, OTBS_0787, OTBS_0946, and OTBS_1546, were primarily composed of extended coiled-coils and long β-strands. For modeling of tertiary structure, we produced 4 models using the SWISS-MODEL server. The best templates were chosen based on GQME and QMEAN4 scores. Model validation was performed using ERRAT, which examines statistical accuracy based on atom interaction patterns within each category of atoms. Overall quality factor ranged between 76% and 100%, which suggests that these are reliable models. Verification of model reliability was also performed using Verify3D, where the average 3D - 1D scores were >0.2 for most residues (Figure 1).

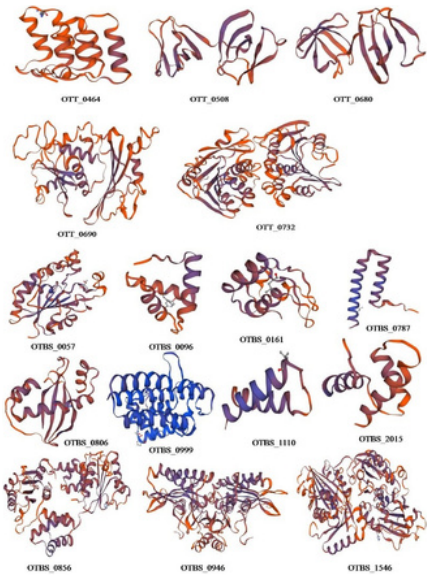


**Figure 1**. Homology modeling of hypothetical proteins.

Additional quality and reliability evaluations using several structural validation tools, including the Ramachandran plot, Z-score, and Q-value, were analyzed. The Z-scores (obtained from the PROSA web server) ranged from −0.93 to −6.5, which agree well with the Z-scores of high-quality homology models. The Ramachandran plot analysis showed that 88–99% of the residues are in favored regions of the plot, and only 0.85–3.3% are in outlier regions, which supports the stereochemical qualities of the models. Taken together, these evaluations, Z-score, Q-value, and Ramachandran plot (Figure 2), determined that the generated homology models were accurate and reliable.

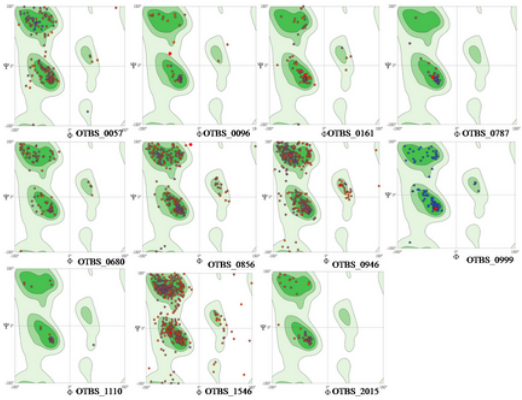


**Figure 2**. Ramachandran plot showing the amino acid residues of HPs present in favored and outlier regions.

## Protein-protein interaction network

To analyze functional protein association networks, we used STRING to predict interactions between each of the hypothetical and conserved proteins and to identify potential partners as well. We included proteins with an interaction score greater than 0.5, and will mention two proteins from the Ikeda strain (OTT_0001 and OTT_0636) with an interaction score of 0.431 and 0.442. Although some proteins act independently, most proteins require association with other proteins in order to exhibit normal biological activity. For this reason, it is important to characterize these protein-protein associations for a proper understanding of protein function and, consequently, cellular biology. The STRING based analysis suggested that some of the hypothetical proteins are likely involved in essential cellular processes- membrane transport, biosynthesis of biomolecules, and translational regulation. The predicted interaction networks also showed the query proteins associate with functionally characterized proteins and some previously uncharacterized ones (Figure 3).

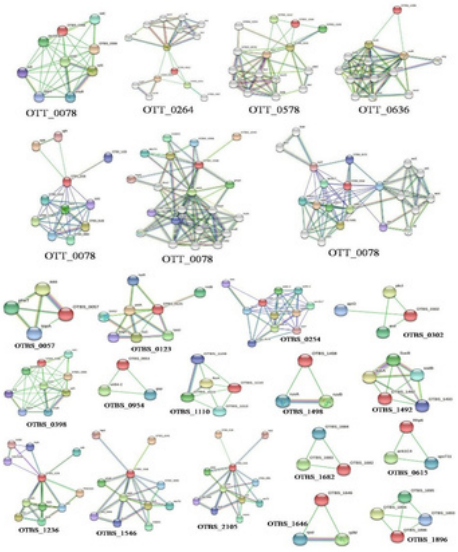


**Figure 3**. Protein-protein interaction of Hypothetical proteins.





https://doi.org/10.61577/imb.2025.100006

## Host non-homology analysis

An ideal drug target should be both essential for the pathogen's survival and pathogen-specific. To avoid undesirable cross-reactivity with host proteins, your target must not have close homologs in the human proteome. With those concepts in mind, we conducted a host of non-homology analyses of hypothetical proteins (HPs) by identifying proteins that have no significant similarity to human proteins. HPs that have been functionally characterized and with high confidence were BLASTp searched against the human proteome (cut-off of e-value 0.0001) and found no substantial matches, suggesting these proteins did not have homologs in humans and could have therapeutic potential.

## Discussions

Essential Genes that are essential for an organism's survival encode proteins whose functions are fundamental to life; therefore, identifying these essential elements provides critical insight into basic biology and reveals high-value targets for therapeutic intervention. To date, experimental characterization of many *O. tsutsugamushi* hypothetical proteins (HPs) is lacking, so we employed an integrative in-silico pipeline to infer their likely functions and prioritize candidates for validation.

Conventional in-silico functional annotation—domain and family detection, subcellular localization, transmembrane helix prediction, and disulfide-bond mapping—offers useful first-pass assignments for HPs. However, mapping protein–protein interactions (PPIs) adds an essential layer of biological context because most cellular processes are executed by multiprotein assemblies rather than isolated polypeptides. Large molecular machines, such as the DNA replication and translation apparatuses, rely on precisely organized PPIs to carry out their roles, and interaction partners often provide strong clues to a protein's biological function [23]. Consequently, PPI data are indispensable for reconstructing pathway membership and cellular networks in the post-genomic era [24].

Evidence from other pathogens underscores the functional and clinical relevance of uncharacterized proteins. Proteomic and bioinformatics studies of drug-resistant Mycobacterium tuberculosis have repeatedly identified the differential expression of hypothetical and uncharacterized proteins. Structural and docking analyses indicate that some small-molecule drugs interact with conserved domains within these unannotated proteins, and their overexpression may contribute to compensatory mechanisms against drug pressure. For example, inducible expression of certain annotated and unannotated proteins (e.g., Rv0148 and Rv3841) increased resistance phenotypes in heterologous systems, suggesting that unknown proteins can modulate drug susceptibility [25]. These observations highlight how careful bioinformatic characterization, including PPI mapping, can predict functional partners and pathways that may underlie resistance or virulence mechanisms, and reveal novel targets for intervention.

From a drug-discovery perspective, an optimal target should be both essential to the pathogen and distinct from host proteins to minimize off-target effects. Database-driven target prioritization (for instance via ChEMBL searches) can rapidly flag druggable candidates; SA0940 and SA0021 were identified as druggable in previous analyses [26], and ribonuclease activity (SA0940) has been proposed as a therapeutic target in Staphylococcus aureus [27]. Applying similar criteria—essentiality, pathogen specificity, structural tractability, and presence of accessible active sites—allowed us to shortlist HPs from *O. tsutsugamushi* that merit focused biochemical and pharmacological follow-up.

Prior computational efforts on *O. tsutsugamushi* provide complementary context: Imam et al. [28] functionally annotated hundreds of HPs in the Karp strain, classifying them into enzymes, transporters, binding proteins, metabolic and catalytic actors, and kinases, and identified a subset predicted to be highly virulent. Their use of sequence-similarity networks to visualize functional trends across superfamilies demonstrated the power of integrative sequence-structure analyses to generate testable hypotheses about protein function. Our study builds on and extends these approaches by combining physicochemical profiling, structural modelling, active-site mapping, host non-homology screening, and PPI inference to produce a prioritized, pathogen-focused set of candidate targets for experimental validation.

In summary, computational annotation and interaction mapping of hypothetical proteins provide a scalable strategy to illuminate uncharacterized regions of the *O. tsutsugamushi* proteome, propose biological roles, and identify pathogen-specific targets for future biochemical assays, structure-based drug design, and investigations into mechanisms of virulence and drug resistance.

## Conclusions

Functional and structural annotation of the hypothetical proteins of *O. tsutsugamushi* strain Ikeda has been done in this study through in-silico methods, representing both direct and indirect links to virulence mechanisms. The results provided the framework to state that these protein domains are important to cell invasion and apoptosis of host cells. The possibility that these protein domains have been identified in many strains of bacteria and that they often interact with multiple proteins further suggests that there may be potential for developing new antibacterial drugs through these protein domains. Studies of protein–protein and protein–ligand docking could elucidate which amino acid residues bind to ligand molecules. Furthermore, molecular docking and simulations to demonstrate stable interactions between hypothetical proteins and all proteins could be beneficial for this study and the further development of new therapeutic agents or vaccines against *O. tsutsugamushi*.

## Acknowledgements

We acknowledge the Medical Research Laboratory of IMS & SUM hospital, Siksha 'O' Anusandhana (deemed to be) University for providing Lab facility.

## Disclosure Statement

The authors declare no conflict of interest.

## Funding

No funding available.

## References

1. Walker DH. Scrub typhus—scientific neglect, ever-widening impact. N Engl J Med. 2016;375(10):913-915. https://doi.org/10.1056/NEJMp1608499
2. Xu G, Walker DH, Jupiter D, Melby PC, Arcari CM. A review of the global epidemiology of scrub typhus. PLoS Negl Trop Dis. 2017;11(11):e0006062. https://doi.org/10.1371/journal.pntd.0006062



https://doi.org/10.61577/imb.2025.100006

3. Mahajan SK, Rolain JM, Kanga A, Raoult D. Scrub typhus involving central nervous system, India, 2004–2006. Emerg Infect Dis. 2010;16(10):1641. https://doi.org/10.3201/eid1610.100456

4. Khan SA, Bora T, Laskar B, Khan AM, Dutta P. Scrub typhus leading to acute encephalitis syndrome, Assam, India. Emerg Infect Dis. 2017;23(1):148. https://doi.org/10.3201/eid2301.161038

5. Jain P, Prakash S, Tripathi PK, Chauhan A, Gupta S, Sharma U, et al. Emergence of *Orientia tsutsugamushi* as an important cause of acute encephalitis syndrome in India. PLoS Negl Trop Dis. 2018;12(3):e0006346. https://doi.org/10.1371/journal.pntd.0006346

6. Panda S, Swain SK, Sarangi R. An epidemiological outbreak of scrub typhus caused by *Orientia tsutsugamushi*—A comprehensive review. J App Biol Biotech. 2022;10:76-83. https://dx.doi.org/10.7324/JABB.2022.100509

7. Luce-Fedrow A, Lehman ML, Kelly DJ, Mullins K, Maina AN, Stewart RL, et al. A review of scrub typhus (*Orientia tsutsugamushi* and related organisms): then, now, and tomorrow. Trop Med Infect Dis. 2018;3(1):8. https://doi.org/10.3390/tropicalmed3010008

8. Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, Fukuhara M, et al. The whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. DNA research. 2008;15(4):185-199. https://doi.org/10.1093/dnares/dsn011

9. Panda S, Swain SK, Sahu BP, Sarangi R. Gene expression and involvement of signaling pathways during host–pathogen interplay in *Orientia tsutsugamushi* infection. 3 Biotech. 2022;12(9):180. https://doi.org/10.1007/s13205-022-03239-7

10. Singh A, Singal B, Nath O, Singh IK. Functional annotation and classification of the hypothetical proteins of Neisseria meningitidis H44/76. Am J Biosci Bioeng. 2015;3:57-64. https://doi.org/10.11648/j.bio.20150305.16

11. Shahbaaz M, ImtaiyazHassan MD, Ahmad F. Functional annotation of conserved hypothetical proteins from Haemophilus influenzae Rd KW20. PloS One. 2013;8(12):e84263. https://doi.org/10.1371/journal.pone.0084263

12. Kolker E, Picone AF, Galperin MY, Romine MF, Higdon R, Makarova KS, et al. Global profiling of Shewanella oneidensis MR-1: expression of hypothetical genes and improved functional annotations. Proc Natl Acad Sci. 2005;102(6):2099-2104. https://doi.org/10.1073/pnas.0409111102

13. Hoskeri J, Krishna V, Amruthavalli C. Functional annotation of conserved hypothetical proteins in Rickettsia massiliae MTU5. Journal of Computer Science & Systems Biology. 2010;3(2):50-51.

14. Gasteiger E, Hoogland C, Gattiker A, Duvaud SE, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server. InThe proteomics protocols handbook 2005; 571-607. https://doi.org/10.1385/1-59259-890-0:571

15. Yu CS, Lin CJ, Hwang JK. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci. 2004;13(5):1402-1406. https://doi.org/10.1110/ps.03479604

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

17. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8(10):785-786. https://doi.org/10.1038/nmeth.1701

18. Falda M, Toppo S, Pescarolo A, Lavezzo E, Di Camillo B, Facchinetti A, et al. Argot2: a large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. BMC Bioinfo. 2012;13(Suppl 4):S14. https://doi.org/10.1186/1471-2105-13-S4-S14

19. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. Appl Crystallogr. 1993;26(2):283-291. https://doi.org/10.1107/S0021889892009944

20. Eisenberg D, Lüthy R, Bowie JU. [20] VERIFY3D: assessment of protein models with three-dimensional profiles. InMethods in enzymology. 1997;277:396-404. https://doi.org/10.1016/S0076-6879(97)77022-8

21. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007;35(suppl_2):W407-10. https://doi.org/10.1093/nar/gkm290

22. Jadhav A, Shanmugham B, Rajendiran A, Pan A. Unraveling novel broad-spectrum antibacterial targets in food and waterborne pathogens using comparative genomics and protein interaction network analysis. Infection, Genet Evol. 2014;27:300-308. https://doi.org/10.1016/j.meegid.2014.08.007

23. Phizicky EM, Fields S. Protein-protein interactions: methods for detection and analysis. Microbiol Rev. 1995;59(1):94-123. https://doi.org/10.1128/mr.59.1.94-123.1995

24. Islam MS, Shahik SM, Sohel M, Patwary NI, Hasan MA. In silico structural and functional annotation of hypothetical proteins of Vibrio cholerae O139. Genomics Inform. 2015;13(2):53. https://doi.org/10.5808/GI.2015.13.2.53

25. Sharma D, Lata M, Faheem M, Khan AU, Joshi B, Venkatesan K, Shukla S, Bisht D. M. tuberculosis ferritin (Rv3841): potential involvement in Amikacin (AK) & Kanamycin (KM) resistance. Biochem Biophys Res Commun. 2016;478(2):908-912. https://doi.org/10.1016/j.bbrc.2016.08.049

26. Prava J, Pranavathiyani G, Pan A. Functional assignment for essential hypothetical proteins of Staphylococcus aureus N315. Int J Biol Macromol. 2018;108:765-774. https://doi.org/10.1016/j.ijbiomac.2017.10.169

27. Redko Y, Galtier E, Arnion H, Darfeuille F, Sismeiro O, Coppee JY, et al. RNase J depletion leads to massive changes in mRNA abundance in Helicobacter pylori. RNA Biol. 2016;13(2):243-253. https://doi.org/10.1080/15476286.2015.1132141

28. Imam N, Alam A, Ali R, Siddiqui MF, Ali S, Malik MZ, Ishrat R. In silico characterization of hypothetical proteins from *Orientia tsutsugamushi* str. Karp uncovers virulence genes. Heliyon. 2019;5(10). https://doi.org/10.1016/j.heliyon.2019.e02734




https://doi.org/10.61577/imb.2025.100006