

## In silico functional and structural characterization revealed virulent proteins of *Francisella tularensis* strain SCHU4

Prerna Goel<sup>1</sup>, Tanya Panchal<sup>1</sup>, Nandini Kaushik<sup>1</sup>, Ritika Chauhan<sup>1</sup>, Sandeep Saini<sup>1,2</sup>, Vartika Ahuja<sup>1</sup>, Chander Jyoti Thakur<sup>1,\*</sup>

1) Department of Bioinformatics, Goswami Ganesh Dutta Sanatan Dharma College, Sector 32 C, Chandigarh, India

2) Department of Biophysics, Panjab University, Sector 25, 160014, Chandigarh, India

### ABSTRACT

*Francisella tularensis* is a pathogenic, aerobic gram-negative coccobacillus bacterium. It is the causative agent of tularemia, a rare infectious disease that can attack skin, lungs, eyes, and lymph nodes. The genome of *F. tularensis* has been sequenced, and ~16% of the proteome is still uncharacterized. Characterizations of these proteins are essential to find new drug targets for better therapeutics. *In silico* characterization of proteins has become an extremely important approach to determine the functionality of proteins as experimental functional elucidation is unable to keep pace with the current growth of the sequence database. Initially, we have annotated 577 Hypothetical Proteins (HPs) of *F. tularensis* strain SCHU4 with seven bioinformatics tools which characterized them based on the family, domain and motif. Out of 577 HPs, 119 HPs were annotated by five or more tools and are further screened to predict their virulence properties, subcellular localization, transmembrane helices as well as physicochemical parameters. VirulentPred predicted 66 HPs out of 119 as virulent. These virulent proteins were annotated to find the interacting partner using STRING, and proteins with high confidence interaction scores were used to predict their 3D structures using Phyre2. The three virulent proteins Q5NH99 (phosphoserine phosphatase), Q5NG42 (Cystathionine beta-synthase) and Q5NG83 (Rrf2-type helix turn helix domain) were predicted to involve in modulation of cytoskeletal and innate immunity of host, H<sub>2</sub>S (hydrogen sulfide) based antibiotic tolerance and nitrite and iron metabolism of bacteria. The above predicted virulent proteins can serve as novel drug targets in the era of antibiotic resistance.

**Keywords:** Domains; Drug target; Virulence factor; Annotation; Microbial resistance; Bioinformatics

### INTRODUCTION

*Francisella tularensis*, which was first accidentally discovered in 1911, is the causative agent of tularemia and is considered a dangerous potential biological weapon. This bacterium is a small facultative, aerobic, non-spore-forming, non-motile intracellular pathogen that mainly replicates in macrophages [1]. The pneumonic form of tularemia is life-threatening and can

\*Corresponding Author: Department of Bioinformatics, Goswami Ganesh Dutta Sanatan Dharma College Sector 32 C, Chandigarh, India, 160030

Tel: +91 8699776533; Fax: +91 1722661077; E. mail: [chander.jyoti@ggdsd.ac.in](mailto:chander.jyoti@ggdsd.ac.in)

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

result in death. According to the Centers for Disease Control and Prevention (CDC), *F.tularensis* is a viable biological warfare agent due to its high infectivity, ability to cause death, and low infectious dose [2]. The family *Francisellaceae* includes three species in the genus *Francisella* and four subspecies of *F. tularensis*. Although all have been associated with human disease, only the *tularensis* and *holarctica* subspecies of *F. tularensis* are relatively common [3]. *F. tularensis* is covered by a capsule that is required for virulence [4,5]. Recently, the presence of Pili (type 4) has been observed which may be the cause of the virulence of the bacteria [6]. *F. tularensis* has been reported in birds, reptiles, fish, invertebrates, and mammals, including humans. It is caused by contact with infected animals or vectors such as ticks, mosquitoes, and deer flies which makes it a dangerous and highly infectious agent for biological warfare [7-10]

The genome of *F. tularensis* strain SCHU4 has been sequenced. Its 1.89 Mb base pair long genome with 1910 total genes is contained in only one chromosome. Approximately 83% of the genome is coding which codes about 1667 proteins. Out of these coding proteins, 577 (~16%) are hypothetical proteins (HPs), i.e., there is no experimental evidence regarding the proteins' function and structure, implying that these proteins are yet to be annotated [11, 12]

Due to the advancements in sequencing technologies, a large amount of protein sequence data is generated that needs to be annotated or characterized but still a significant proportion is demarcated as hypothetical. To keep pace with the exponential growth of sequence data, computational approaches play an important role in deciphering the function of HPs that reduce the laborious experimental protocols [13, 14]. These approaches are important for understanding species-specific pathogenicity, new target identification, pathway associations, and evolutionary processes [15-17].

In-silico approach has been successfully applied previously to various species like *Mycobacterium tuberculosis*, *Streptococcus Gordonii*, *Vibrio cholera*, *Staphylococcus aureus*, *Haemophilus influenza*, *Shigella flexneri*, *Exiguobacterium antarcticum* strain B7, *Neisseria gonorrhoeae* and *Chloroflexus aurantiacs* [18-26].

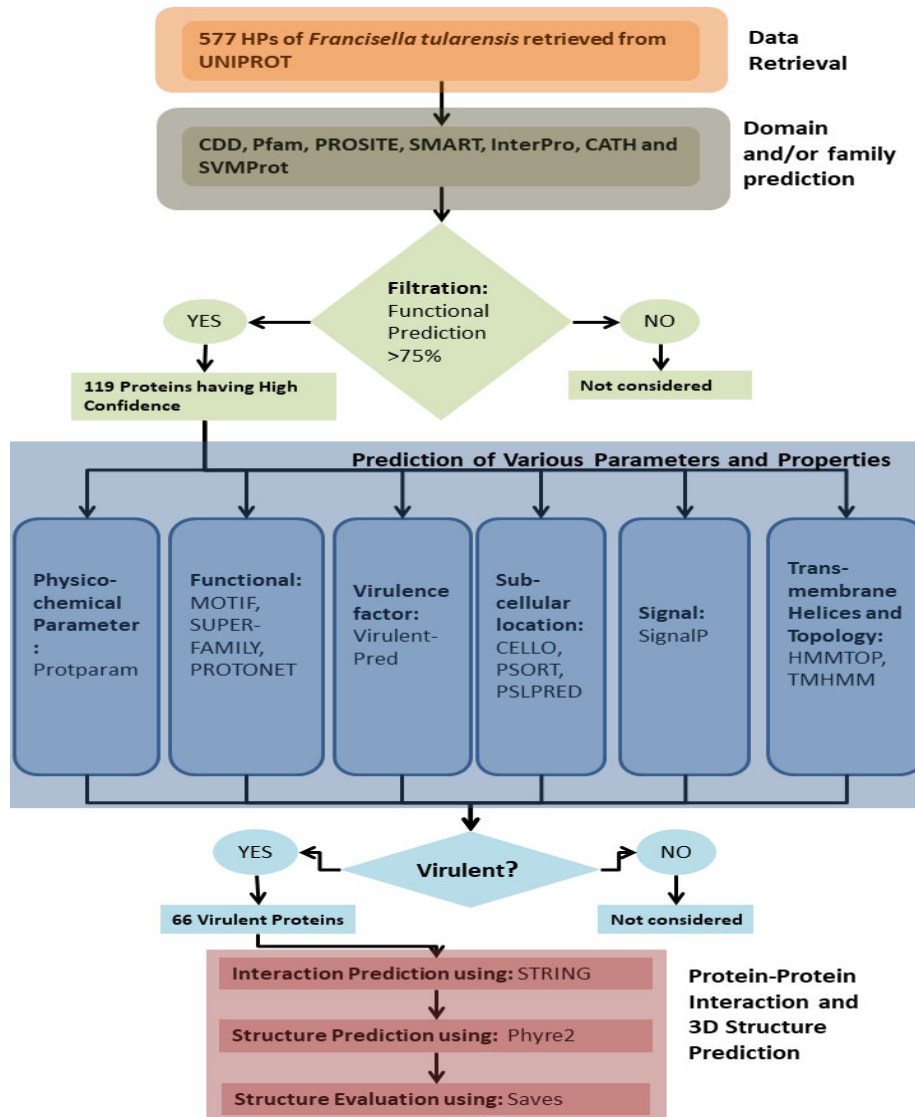
Hence, considering the bacterium's pathogenicity and the need to discover a new drug target in the era of antimicrobial resistance, we have computationally annotated uncharacterized proteins of *F. tularensis* strain SCHU4 using various bioinformatics tools. The study also identified the most virulent proteins from 577 HPs that can serve as novel drug targets.

## MATERIALS AND METHODS

**Data Retrieval:** In this study, hypothetical proteins of *F. tularensis* strain SCHU4 were retrieved from UniprotKB (<https://www.uniprot.org/>) [27]. Out of the total 3,733 proteins of *F. tularensis*, 577 were uncharacterized proteins. The computational workflow used for this study has been illustrated in Figure 1.

**Functional annotation and filtration of HPs:** To assign biological function to HPs is a major obstacle and a challenging task, we have used various databases and tools for functional annotation including CDD (Conserved Domain Database) (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) [28], SMART (Simple Modular Architecture Research Tool) (<http://smart.embl-heidelberg.de/>) [29] CATH (Class, Architecture, Topology, Homology) (<https://www.cathdb.info/>) [30], PROSITE (<https://prosite.expasy.org/>) [31], Pfam (<http://pfam.xfam.org/>) [32], InterPro (<https://www.ebi.ac.uk/interpro/>) [33] and SVMProt (<https://bio.tools/svm-prot>) [34]. CDD, SMART, and PROSITE were used for identification of conserved domains, and CATH was used to classify the domains within a structural hierarchy. Pfam, SVMProt, and InterPro were used to classify the HPs into functional families based on similarity. We have selected only those HPs for further analyses, which were annotated by five or more tools out of seven (i.e.75% annotation criteria.). To validate our initial annotation result we have further used MOTIF (<https://www.genome.jp/tools/motif/>) [35], SUPERFAMILY

(<https://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/>) [36], PROTONET (<http://www.protonet.cs.huji.ac.il/>) [37] to classify the HPs into functional families. The MOTIF assigns function to the HPs by aligning with the profile generated from multiple sequence alignment while ProtoNet is based automated hierarchical clustering that rely on all against all BLAST (Basic Local Alignment Search Tool) similar search and map HPs into an existing clustering and SUPERFAMILY predict domain in HPs based Structural Classification of Proteins (SCOP) database hierarchy at superfamily level that groups together the most distantly related domains [35-37]. Details of analysis were given in supplementary Tables S1 and S2.



**Figure 1:** Flowchart of Methodology

**Prediction of physicochemical parameters and subcellular location:** For physiochemical parameters prediction ProtParam (<https://web.expasy.org/protparam/>) was used. We retrieved various physicochemical properties of the protein parameters i.e. .molecular mass, theoretical isoelectric point (pI), amino acid composition, atomic composition, extinction coefficient, instability index, aliphatic index, and grand average hydropathy (GRAVY) [38].

To determine the subcellular localization we used three predictors; CELLO, PSLpred, and PSORTb. CELLO (subcellular LOcalization predictor) (<http://cello.life.nctu.edu.tw/>), a two-

level support vector machine to predict the subcellular locations [39]. On the other hand, PSLpred (<https://webs.iitd.edu.in/raghava/pslpred/submit.html>) relies on SVM method [40] and PSORTb (<https://www.psort.org/psortb/>) combine experimental and computational prediction for prediction of subcellular locations of HPs [41]. TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) [42], HMMTOP (<http://www.enzim.hu/hmmtop/html/submit.html>) [43] and SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) [44] were used to determine the transmembrane helices and signal peptide in HPs.

**Prediction of virulence factor:** Virulence factor analysis was done by VirulentPred (<http://203.92.44.117/virulent/submit.html>), a SVM-based tool that calculates potential virulent proteins and use five-fold cross-validation technique. We have selected default threshold value i.e. 0.0 and score from 0.0 to 1 predict the protein as virulent and minus score indicate non-virulent protein [45].

**Protein-protein interaction prediction and tertiary structure prediction:** STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) (<https://string-db.org/>) database was used to predict the protein-protein interactions of the virulent HPs [46]. Among 66 HPs, we have selected only those proteins which were associated with other known proteins with high confidence score. STRING provides a functional link between proteins called networks along with their scores.

Furthermore, we have determined the 3D structure of those proteins that possess interactions with known proteins with a high confidence score. To predict the protein structure, which is more conserved than protein sequence, we use a protein homology/analogy recognition engine (Phyre 2) (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) that gather homologous sequences against query protein and then scanned against HMM database of known protein structure [47, 48]. The structure evaluation was performed by Procheck using SAVES v6.0 server (<https://saves.mbi.ucla.edu/>) that analyzes the number of residues falling in the allowed region in the Ramachandran Plot as a measure of quality [49].

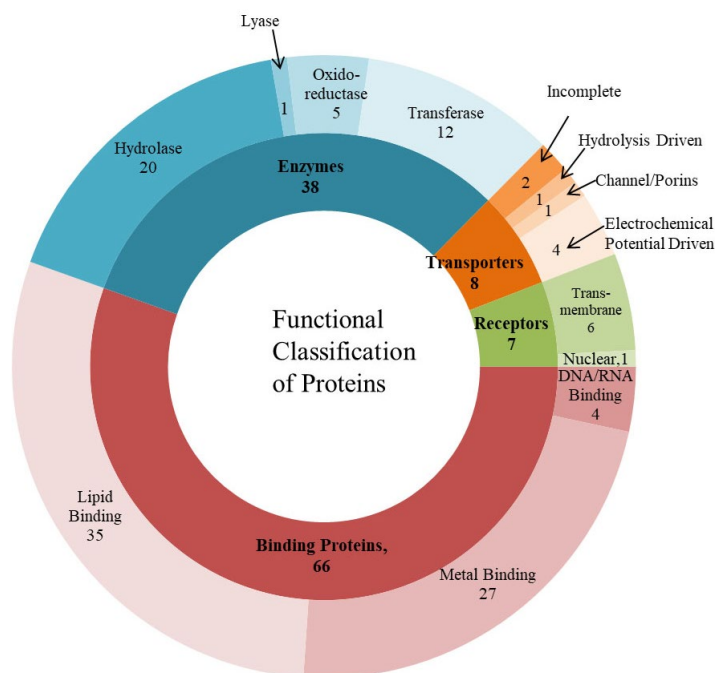
## RESULTS

We have extensively analyzed 577 HPs of *F. tularensis* using various bioinformatics tools, i.e., CDD, Pfam, SMART, PROSITE, InterPro, CATH, and SVMProt, for domain and/or functional prediction. If the same function is assigned for HPs by five or more tools, then only those HPs were selected for further annotation. Using this strategy, we have identified 119 HPs out of 577. These 119 HPs are further analyzed by using Motif, Superfamily, and Protonet for protein family characterization and were classified as 38 enzymes, 66 binding proteins, 7 receptors, 8 transport proteins as given in Fig. 2 and supplementary Table S2.

Protoparam was used for the prediction of various physicochemical parameters of the filtered 119 proteins. These parameters were molecular weight, aliphatic index, instability index, theoretical pI, etc. >40. The instability index of the protein signifies that the protein is unstable. We identified 28 proteins whose instability index values were greater than 40. The theoretical pI of the HPs lies between 10 and 4. The GRAVY values indicate both hydrophobic (negative) and hydrophilic (positive) HPs [50-52]. The Aliphatic Index of the HPs varied between 160 and 59. The physicochemical parameters of 119 HPs are listed in supplementary Table S3.

Subcellular location correlates protein biological function as cell offer different chemical environment and interaction [53]. Signal peptides control transport of proteins to target location in all organisms [54, 55]. Membrane proteins are involve in signaling between cells, transport and energy transduction [42, 43]. Based on the average subcellular prediction results of the three tools HPs were located cytoplasm (40%), membrane (39%), extracellular region (18%), and periplasm (3%). We have predicted presence of signal peptide sequence in 78 HPs. HMMTOP and TMHMM server predicted 65 and 61 proteins as transmembrane helices. These prediction

results help to better understanding the new drug target and detail results are mentioned in supplementary Table S4.



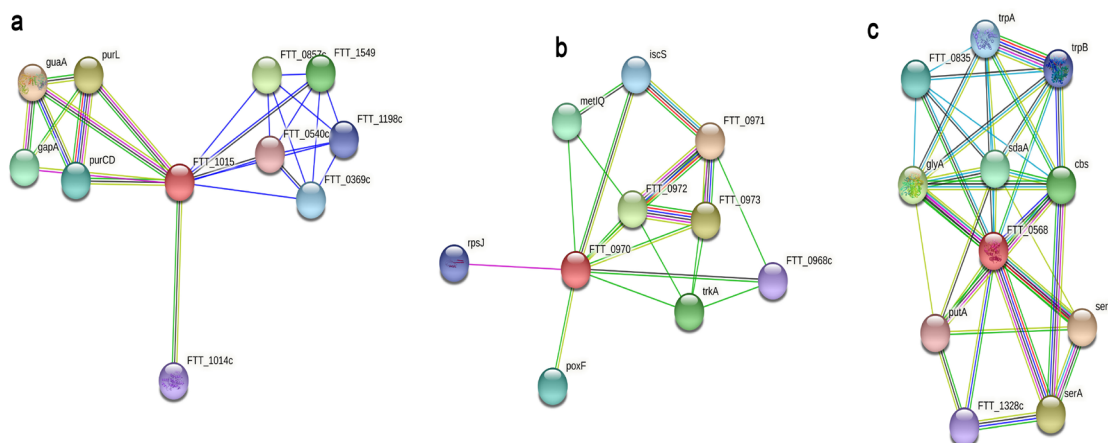
**Figure 2:** 119 HPs classified into enzymes (38), binding proteins (66), receptors (7) and transport proteins (8).

Prediction of virulent protein is important for identification of novel targets and understanding the complex virulence mechanism of pathogen [45]. VirulentPred predicted 66 HPs as virulent with the score ranges from 0.95 to 1 which is a higher score that minimize false positive hits (supplementary Table S5).

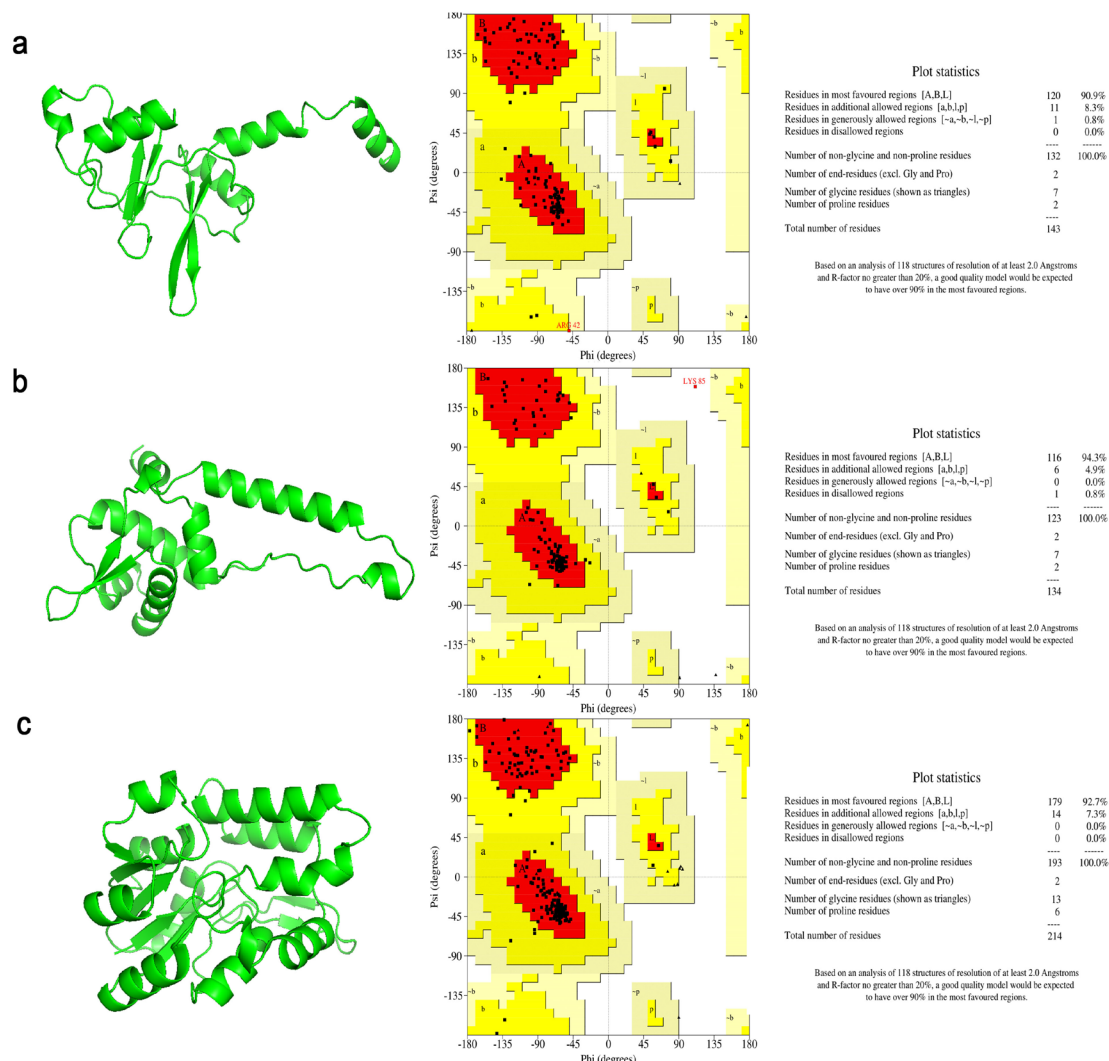
Protein-protein interaction network: We have predicted the association of 66 HPs with other proteins that can further elaborate the role of HPs. Among the complete set of 66 proteins, 25 were dropped out as these did not show any significant interactions. The remaining 41 proteins are related to the known genes with high confidence. We identified interacting partners involved in lipid biosynthesis, signal transduction, transcriptional regulators, iron sulfur binding, extracellular signals, and stimuli. Some proteins interacted with the type 4 pilli, which is suspected to be the cause of the pathogenicity of *F. tularensis* [6]. Our study found that out of 41 proteins, three proteins Q5NH99, Q5NG42, and Q5NG83, had a very high confidence score of 98%, 95%, and 89.5%, respectively, with their interacting partner. Q5NG83 interacts with the ABC Transporter enzyme, which plays a role in bacterial virulence mainly associated with the uptake of nutrients, while Q5NG42 interacts with GMP synthase required for virulence factor production and Q5NH99 interact with proteins involved in metabolism [56, 57]. The STRING results are shown in Figure 3.

Phyre2, a web-based server to predict protein structure based on the principle of homology-based modeling. Query protein is scanned against HMMs proteins of known structure, and top-scoring alignments are used to construct crude backbone, and then model is corrected by loop modeling, and side-chain is added to generate the final model [47, 48]. We have successfully generated the three best models with high confidence i.e., Q5NG42 protein with 99.9% coverage and 70% confidence, Q5NG83 with 100% confidence and 98% coverage, Q5NH99 with 100% confidence and coverage. The structures predicted by using Phyre2 were evaluated by Procheck to check the quality of the 3D model and more than 90% residues were in favoured region of Ramachandran Plot for all the three predicted structure as shown in Figure 4.





**Figure 3:** Protein-protein Interaction Network of Q5NG42, Q5NG83 and Q5NH99 proteins by STRING database



**Figure 4:** Three-Dimensional structural prediction of (a) Q5NG42, (b) Q5NG83 and (c) Q5NH99 proteins by Phyre2 and evaluation of structure by Ramachandran plot by SAVES which show score above 90%.

## DISCUSSION

Cost-effective sequencing technology has made bacterial genome sequencing affordable to several labs world widely [58]. The published bacterial genome consists of a large group of hypothetical proteins which must be annotated to understand disease pathogenesis, virulent determinants, survival, and propagation. Characterization of HPs by various bioinformatics tools may identify potential biomarkers for the screening and diagnostics of diseases as well as pharmacological targets to design and discover novel drugs [59].

In the present study, we have annotated HPs of *F. tularensis* by various computational tools to predict the function and structure of proteins that can potentially be involved in the toxicity of the pathogen. The initial step involved identification of conserved region i.e. domain and motif which provide the clue about protein function. We started analysis of 577 HPs using seven functional tools and those HPs are selected for further analysis if the result of each HPs is predicted by five or more tools i.e. >75% prediction analysis which retrieves 119 HPs. These HPs were further annotated for various properties including physicochemical parameters like GRAVY index, aliphatic index, pI, Instability index, cellular localization, transmembrane analysis, functional analysis, and virulent prediction. We then predicted interacting partners of these 66 proteins and based on high confidence score we have selected 3 virulent proteins Q5NH99, Q5NG42 and Q5NG83 with 98%, 95% and 89.5% confidence score respectively (Table 1).

**Q5NH99** predicted as phosphoserine phosphatase (PSP), a key essential metabolic enzyme catalyse the dephosphorylation of phosphoserine to serine and inorganic phosphate and is a member of haloacid dehalogenase (HAD) superfamily. PSP have been demonstrated to facilitate entry of *Porphyromonas gingivalis* into host cells by modulating host cytoskeletal architecture, innate immune responses, dephosphorylating colicin and NF- $\kappa$  [60-62].

**Table 1: Functional and structural annotation results of virulent proteins:** Three virulent proteins with their annotated function, virulence score (0.4 is considered as threshold), and Phyre2 structure prediction and evaluation score.

Protein ID	Annotated Function	VirulentPred Score	Final Prediction	Structure Modelling confidence score by Phyre2	Structure Evaluation (percent amino acid residues in most favoured region)
Q5NH99	Phosphoserine phosphatase	1.0016	Virulent	98%	92.7%
Q5NG42	Cystathionine beta-synthase	0.9776	Virulent	95%	90.9%
Q5NG83	Rrf2-type helix turn helix domain	1.0957	Virulent	89.5%	94.3%

**Q5NG42**, a 196 amino acid length was predicted as Cystathioninebeta-synthase (CBS) catalyzes a condensation reaction to convert homocysteine to cystathionine. This enzyme also produces hydrogen sulfide (H<sub>2</sub>S) during this catalysis reaction [63]. A number of previous studies in bacterial species have revealed that production of H<sub>2</sub>S provides antibiotic tolerance and serve as universal defence mechanism [66-66]. Further supportive evidence was gained by study in which the inactivation of CBS leads to decrease in H<sub>2</sub>S level and hence more sensitivity towards antibiotics [67]. Hence, CBS can act as a main target against antibiotic resistance in era of antimicrobial resistance (AMR).

**Q5NG83** Rrf2-type HTH domain is a 130 residues long DNA binding, helix-turn-helix domain. This domain is present in the proteins that serve as transcription regulators of rrf2

family. Usually, these transcription regulators regulate the genes of bacteria that are involved in nitrite and iron metabolism. For instance, in *Escherichia coli* one the regulator protein, nsR that contains Rrf2-type HTH domain regulates the NO (nitric oxide) sensitive transcription and thus protects the cell from the NO stress. Similar transcription regulator protein with same role was also found in other bacteria such as *Bacillus subtilis* and *Streptomyces coelicolor* whereas *Rhizobium leguminosarum* contains a transcription regulator protein rirA which is an iron-responsive regulator that involved in uptake of iron [68, 69].

Our study has predicted virulent proteins that can play a vital role in future therapeutics of tularemia. These results need to be analyzed in vitro to verify these results experimentally to check the feasibility of these proteins as possible drug targets and their feasibility. These proteins can be essential in determining the virulence mechanism as well as act as the keys for unlocking the mystery related to the immunity of *F. tularensis* to antibiotics. By further analysis of these proteins, we can also look for binding sites and try to dock leads to these proteins to develop a treatment that has a higher efficacy with fewer side effects.

A recent study by Kumar et al. used pan-genome analysis and genome mining approach to explore the diversity of the genus Francisella. The whole-genome sequence comparisons of the 26 genomes in their study revealed average nucleotide identity (ANI) of 97.0–99.5% for the human pathogens strains cluster, whereas ANI values of 74.2–90.4% were found in the strains that are mainly related to water and environment [70].

Thus the above study shows that there exhibit a good degree of similarity in the coding regions of Francisella genomes. Hence, the functional characterization of HPs in our study may help in the study of proteins of other strains of Francisella

**Acknowledgements:** The authors are thankful to University Grant Commission (UGC) and Department of Science & Technology (DST), India for providing facilities.

**Conflict of Interest:** Authors have no conflict of interest.

## REFERENCES

1. Hirschmann JV. From squirrels to biological weapons: The early history of Tularemia. *Am J Med Sci* 2018;356:319-328.
2. Dennis DT, Inglesby TV, Henderson DA, Bartlett JG, Ascher MS, Eitzen E, Fine AD, Friedlander AM, Hauer J, Layton M, Lillibridge SR, McDade JE, Osterholm MT, O'Toole T, Parker G, Perl TM, Russell PK, Tonat K. Working Group on Civilian Biodefense. Tularemia as a biological weapon: medical and public health management. *JAMA* 2001; 285:2763-2773.
3. Sjöstedt A. Tularemia: history, epidemiology, pathogen physiology, and clinical manifestations. *Ann N Y Acad Sci* 2007;1105:1-29.
4. Sandström G, Löfgren S, Tärnvik A. A capsule-deficient mutant of Francisella tularensis LVS exhibits enhanced sensitivity to killing by serum but diminished sensitivity to killing by polymorphonuclear leukocytes. *Infect Immun* 1988;56:1194-1202.
5. Okan NA, Chalabaev S, Kim TH, Fink A, Ross RA, Kasper DL. Kdo hydrolase is required for Francisella tularensis virulence and evasion of TLR2-mediated innate immunity. *mBio* 2013;4:e00638-12.
6. Salomonsson EN, Forslund AL, Forsberg A. Type IV Pili in Francisella - A virulence trait in an intracellular pathogen. *Front Microbiol* 2011;2:29.
7. Chomel BB, Morton JA, Kasten RW, Chang CC. First pediatric case of Tularemia after a Coyote Bite. *Case Rep Infect Dis* 2016;2016:8095138.
8. Akimana C, Kwaik YA. Francisella-arthropod vector interaction and its role in patho-adaptation to infect mammals. *Front Microbiol* 2011;2:34.



9. Gunnell MK, Adams BJ, Robison RA. The genetic diversity and evolution of *Francisella tularensis* with comments on detection by PCR. *Curr Issues Mol Biol* 2016;18:79-91
10. McLendon MK, Apicella MA, Allen LA. *Francisella tularensis*: taxonomy, genetics, and Immunopathogenesis of a potential agent of biowarfare. *Annu Rev Microbiol* 2006;60:167-185.
11. Larsson P, Oyston PC, Chain P, Chu MC, Duffield M, Fuxelius HH, Garcia E, Hälltorp G, Johansson D, Isherwood KE, Karp PD, Larsson E, Liu Y, Michell S, Prior J, Prior R, Malfatti S, Sjöstedt A, Svensson K, Thompson N, Vergez L, Wagg JK, Wren BW, Lindler LE, Andersson SG, Forsman M, Titball RW. The complete genome sequence of *Francisella tularensis*, the causative agent of tularemia. *Nat Genet* 2005;37:153-159.
12. Lubec G, Afjehi-Sadat L, Yang JW, John JP. Searching for hypothetical proteins: theory and practice based upon original data and literature. *Prog Neurobiol* 2005;77:90-127.
13. Bock JR, Gough DA. Predicting protein-protein interactions from primary structure. *Joel. Bioinformatics* 2001;17:455-460.
14. Doerks T, Von Mering C, Bork P. Functional clues for hypothetical proteins based on genomic context analysis in prokaryotes. *Nucleic Acids Res* 2004;32:6321-6326.
15. Adams MA, Suits MDL, Zheng J, Jia Z. Piecing together the structure-function puzzle: Experiences in structure-based functional annotation of hypothetical proteins. *Proteomics* 2007;7:2920-2932.
16. Nourani E, Khunjush F, Durmus S. Computational approaches for prediction of pathogen-host protein-protein interactions. *Front Microbiol* 2015;6:1-10.
17. Windsor AJ, Mitchell-Olds T. Comparative genomics as a tool for gene discovery. *Curr Opin Biotechnol* 2006;17:161-167.
18. Lee BY, Hefta SA, Brennan PJ. Characterization of the major membrane protein of virulent *Mycobacterium tuberculosis*. *Infect Immun* 1992;60:2066-2074.
19. Telkar S, Kumar H, Mahmood R. Synteny approach of drug target prediction among unique hypothetical proteins of *Streptococcus Gordonii* causing infective endocarditis. *Sci Technol Arts Res J* 2014;2:34.
20. 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;13:53-59.
21. School K, Marklevitz J, Schram WK, Harris LK. Predictive characterization of hypothetical proteins in *Staphylococcus aureus* NCTC 8325. *Bioinformatics* 2016;12:209-220.
22. Shahbaaz M, Ahmad F, Imtaiyaz Hassan M. Structure-based functional annotation of putative conserved proteins having lyase activity from *Haemophilus influenzae*. *3 Biotech* 2015;5:317-336.
23. Gazi MA, Mahmud S, Fahim SM, Kibria MG, Palit P, Islam MR, Rashid H, Das S, Mahfuz M, Ahmeed T. Functional prediction of hypothetical proteins from *Shigella flexneri* and validation of the predicted models by using ROC curve analysis. *Genomics Inform* 2018;16:e26.
24. da Costa WLO, Araújo CLA, Dias LM, Pereira LCS, Alves JTC, Araújo FA, Folador EL, Henriques I, Silva A, Folador ARC. Functional annotation of hypothetical proteins from the *Exiguobacterium antarcticum* strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. *PLoS One* 2016;13:e0198965.
25. Singh A, Singal B, Nath O, Singh IK. Functional annotation and classification of the hypothetical proteins of *Neisseria meningitidis* H44/76. *AJBIO* 2015;3:57-64.
26. Thakur CJ, Saini S, Notra A, Chauhan B, Arya S, Gupta R, Thakur J, Kumar V. Deciphering the functional role of hypothetical proteins from *Chloroflexus aurantiacs* J-10-f1 using bioinformatics approach. *Mol Biol Res Commun* 2020;9:129-139.
27. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* 2004;32:D115-D119.

28. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* 2020;48:D265-D268.
29. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res* 2000;28:231-234.
30. Orengo CA, Pearl FM, Bray JE, Todd AE, Martin AC, Lo Conte L, Thornton JM. The CATH Database provides insights into protein structure/function relationships. *Nucleic Acids Res* 1999;27:275-279.
31. Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJ. The PROSITE database. *Nucleic Acids Res* 2006;34:D227-D230.
32. Bateman A, Birney E, Durbin R, Eddy SR, Finn RD, Sonnhammer ELL. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res* 1999;27:260-262.
33. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F, Sigrist CJ, Zdobnov EM. InterPro--an integrated documentation resource for protein families, domains and functional sites. *Bioinformatics* 2000;16:1145-1150.
34. Li YH, Xu JY, Tao L, Li XF, Li S, Zeng X, Chen SY, Zhang P, Qin C, Zhang C, Chen Z, Zhu F, Chen YZ. SVM-Prot 2016: A web-server for machine learning prediction of protein functional families from sequence irrespective of similarity. *PLoS One* 2016;11:e0155290.
35. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. *Nucleic Acids Res* 2002;30:42-46.
36. Wilson D, Madera M, Vogel C, Chothia C, Gough J. The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Res* 2007;35:D308-D313.
37. Sasson O, Vaaknin A, Fleischer H, Portugaly E, Bilu Y, Linial N, Linial M. ProtoNet: hierarchical classification of the protein space. *Nucleic Acids Res* 2003;31:348-352.
38. Garg VK, Avashthi H, Tiwari A, Jain PA, Ramkete PW, Kayastha AM, Singh VK. MFPPi – Multi FASTA ProtParam Interface. *Bioinformatics* 2016;12:74-77.
39. Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, Lu CH. CELLO2GO: a web server for protein subcellular Localization prediction with functional gene ontology annotation. *PLoS One* 2014;9:e99368.
40. Bhasin M, Garg A, Raghava GPS. PSLpred: Prediction of subcellular localization of bacterial proteins. *Bioinformatics* 2005;21:2522-2524.
41. Rey S, Acab M, Gardy JL, Laird MR, deFays K, Lambert C, Brinkman FS. PSORTdb: a protein subcellular localization database for bacteria. *Nucleic Acids Res* 2005;33:D164-D168.
42. Chen Y, Yu P, Luo J, Jiang Y. Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. *Mamm Genome* 2003;14:859-865.
43. Tusnady GE, Simon I. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol* 1998;283:489-506.
44. Bendtsen JD, Nielsen H, Von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004;340:783-795.
45. Garg A, Gupta D. VirulentPred: A SVM based prediction method for virulent proteins in bacterial pathogens. *BMC Bioinformatics* 2008;9:1-12.
46. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 2017;45:D362-D368.
47. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 2015;10:845-858.

48. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 2009;4:363-371.
49. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993;26:283-291.
50. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker. *The Proteomics Protocols Handbook*. Totowa : Humana Press 2005;571-607.
51. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 1999;112:531-552.
52. Behbahani M, Rabiei P, Mohabatkari H. Analysis and comparison of physiochemical properties, mutations and glycosylation patterns between RNA polymerase and membrane protein of SARS-CoV and SARS-CoV-2. *Mol Biol Res Commun* 2021;10:171-178.
53. 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;13:1402-1406.
54. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 1997;10:1-6
55. Lumangtad LA, Bell TW. The signal peptide as a new target for drug design. *Bioorg Med Chem Lett* 2020;30:127115.
56. Chitty JL, Tatzenko TL, Williams SJ, Koh YQ, Corfield EC, Butler MS, Robertson AA, Cooper MA, Kappler U, Kobe B, Fraser JA. GMP Synthase Is Required for Virulence Factor Production and Infection by *Cryptococcus neoformans*. *J Biol Chem* 2017;292:3049-3059.
57. Garmory HS, Titball RW. ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies. *Infect Immun* 2004;72:6757-6763.
58. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *J Appl Genet* 2011;52:413-435.
59. Hawkins T, Kihara D. Function prediction of uncharacterized proteins. *J Bioinform Comput Biol* 2007;5:1-30.
60. Arora G, Tiwari P, Mandal RS, Gupta A, Sharma D, Saha S, Singh R. High throughput screen identifies small molecule inhibitors specific for *Mycobacterium tuberculosis* phosphoserine phosphatase. *J Biol Chem* 2014;289:25149-25165.
61. Moffatt CE, Inaba H, Hirano T, Lamont RJ. *Porphyromonas gingivalis* SerB-mediated dephosphorylation of host cell cofilin modulates invasion efficiency. *Cell Microbiol* 2012;14:577-588.
62. Tribble GD, Mao S, James CE, Lamont RJ. A *Porphyromonas gingivalis* haloacid dehalogenase family phosphatase interacts with human phosphoproteins and is important for invasion. *Proc Natl Acad Sci U S A* 2006;103:11027-11032.
63. Zhu H, Blake S, Chan KT, Pearson RB, Kang J. Cystathionine  $\beta$ -synthase in physiology and cancer. *Biomed Res Int* 2018;28:3205125.
64. Kimura H. Production and physiological effects of hydrogen sulfide. *Antioxid Redox Signal* 2014;20:783-793.
65. Ng SY, Ong KX, Surendran ST, Sinha A, Lai JJH, Chen J, Liang J, Tay LKS, Cui L, Loo HL, Ho P, Han J, Moreira W. Hydrogen sulfide sensitizes *acinetobacter baumannii* to killing by antibiotics. *Front Microbiol* 2020;11:1875.
66. Shukla P, Khodade VS, SharathChandra M, Chauhan P, Mishra S, Siddaramappa S, Pradeep BE, Singh A, Chakrapani H. "On demand" redox buffering by H<sub>2</sub>S contributes to antibiotic resistance revealed by a bacteria-specific H<sub>2</sub>S donor. *Chem Sci* 2017;8:4967-4972.
67. Shatalin K, Shatalina E, Mironov A, Nudler E. H<sub>2</sub>S: a universal defense against antibiotics in bacteria. *Science* 2011;334:986-990.
68. Bodenmiller DM, Spiro S. The yjeB (nsrR) gene of *Escherichia coli* encodes a nitric oxide-sensitive transcriptional regulator. *J Bacteriol* 2006;188:874-881.

69. Keon RG, Fu R, Voordouw G. Deletion of two downstream genes alters expression of the hmc operon of *Desulfovibrio vulgaris* subsp. *vulgaris* Hildenborough. *Arch Microbiol* 1997;167:376-383.
70. Kumar R, Broms JE, Sjostedt A. Exploring the diversity within the genus *Francisella*-An integrated Pan-Genome and Genome-Mining approach. *Front Microbiol* 2020;11:1928.