

Structural and Functional Annotation of Hypothetical Protein of *Fusobacterium nucleatum* Strain MJR7757B: An in Silico Approach

Md. Isrifil Hossen^{1,2}, Fouzia Mostafa³, Nusrat Jahan², Jannatul Ferdous⁴, Amgad Albahi^{4,5}, Sayed Mashequl Bari^{6*}

¹College of Food Science and Technology, Huazhong Agricultural University, Wuhan, China

²Department of Crop Science and Technology, Rajshahi University, Rajshahi, Bangladesh

³Abdul Malek Ukil Medical College, Noakhali, Bangladesh

⁴Department of Medicine, IBN Sina Medical College, Dhaka, Bangladesh

⁵National Food Research Centre, Khartoum, Sudan

⁶Department of Aquatic Animal Health Management, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh

Email: isrifil.crp@gmail.com, fouziamostafa47@gmail.com, fowara329@gmail.com, janpoly23@gmail.com, amgadelbahi@hotmail.com, *bari.aahm@sau.edu.bd

How to cite this paper: Hossen, Md.I., Mostafa, F., Jahan, N., Ferdous, J., Albahi, A. and Bari, S.M. (2024) Structural and Functional Annotation of Hypothetical Protein of *Fusobacterium nucleatum* Strain MJR7757B: An in Silico Approach. *Computational Molecular Bioscience*, 14, 17-33.

<https://doi.org/10.4236/cmb.2024.141002>

Received: November 7, 2023

Accepted: February 16, 2024

Published: February 19, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Fusobacterium nucleatum is an anaerobic, commensal, gram-negative oral bacterium that is carcinogenic and causes a wide range of human diseases. The present study focused on the analysis of the hypothetical protein, HMPREF3221_01179, derived from *F. nucleatum* strain MJR7757B, employing various computational methods to anticipate both its structure and functional characteristics. NCBI conserved domain analysis, NCBI BLASTp and MEGA Phylogenetic tree study characterize the target protein as an outer membrane efflux protein (ToIC family) which facilitate the bacterial trans-membrane transport. With a molecular weight of 52120.02 Da, an isoelectric point (pI) of 8.33, and an instability index of 29.47, the protein is anticipated to exhibit good solubility in the extracellular space and crucial stability for pharmaceutical applications. The protein's structure meets quality standards during the construction and refinement of its 3D model. The efflux inhibitor Arginine beta-naphthylamide exhibits a significant binding affinity (-7.1 kcal/mol) to the binding site of the target protein. The in-silico analysis improves the understanding of the protein and facilitates future investigations into therapeutic medication.

Keywords

Fusobacterium nucleatum, In Silico, Bacteria, Hypothetical Protein, Molecular Docking

1. Introduction

Fusobacterium nucleatum is a prevalent bacterium found in the mouth that has been associated with several human diseases, such as the formation and advancement of colorectal cancer (CRC) [1]. *F. nucleatum* triggers inflammation, which causes genetic instability and inhibits the body's immune responses against tumors [2] [3]. This gram-negative anaerobic species also associated with adverse pregnancy outcomes, gastrointestinal disorders, cardiovascular disease, rheumatoid arthritis, respiratory tract infections, Lemierre's syndrome, and Alzheimer's disease [4] [5] [6] [7] [8]. *F. nucleatum* infections commonly respond well to standard antibiotic therapies. Among the effective antibiotics are metronidazole, clindamycin, and beta-lactam antibiotics such as penicillin or amoxicillin [9] [10]. Despite advancements in genomic sequencing, a substantial portion of *F. nucleatum*'s proteome remains uncharacterized, including the hypothetical protein HMPREF3221_01179.

Hypothetical proteins (HPs), present in genomes, lack experimental characterization yet are essential for diverse cellular processes and signaling pathways. Their annotation is crucial for comprehending disease mechanisms, aiding drug design, vaccine production, and identifying virulent proteins in bacteria through in-silico studies, offering valuable insights into diseases and pathogenesis [11]. In the field of bioinformatics, researchers are actively unveiling the biological functions and characteristics of millions of uncharacterized proteins from different organisms, which perform a wide range of functions, including structuring cells and organisms and participating in vital *in vivo* processes through interactions with other molecules [12] [13]. By employing bioinformatics methods, researchers can analyze protein structures in 3D, identify new domains, and uncover the functions of proteins, enhancing our understanding of their biological roles [14]. In cases where experimental determination of a protein's function is challenging, function inference can be achieved through sequence similarity; if this fails, analysis of protein structure offers valuable functional clues, with recent advancements in combining various structure-based approaches and integrating evidence from multiple sources [15] [16] [17]. Understanding the role of such proteins is pivotal for comprehending the pathogenicity and biology of this bacterium.

This study focused on the hypothetical protein HMPREF3221_01179 from *F. nucleatum*, a bacterium associated with diverse human infections. Using in silico methods, we have investigated the structural and functional annotations of the hypothetical protein (accession no. KXA20922.1) from the *F. nucleatum* strain MJR7757B.

2. Materials and Methods

2.1. Hypothetical Protein Sequence Retrieval

There are over 400 genome sequences of *F. nucleatum* accessible in the National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov>) [18]. This research select a hypothetical protein HMPREF3221_01179 (accession no. KXA20922.1) from the *F. nucleatum* strain MJR7757B. This protein consists of 438 amino acid residues, and its primary sequence was retrieved in FASTA format for in-depth analysis [19].

2.2. Analysis of Physicochemical Properties of Hypothetical Protein

The physical and chemical properties of the target hypothetical protein were analyzed using the ProtParam tool available on the ExPASy website (<http://web.expasy.org/protparam/>) [20]. These properties included molecular weight, aliphatic index (AI) [21], extinction coefficients [22], GRAVY (grand average of hydropathy) [21], and isoelectric point (pI) [23].

2.3. Hypothetical Protein (Conserved Domains) Function Prediction

The conserved domain analysis of the hypothetical protein was conducted using NCBI Conserved Domain Search Service (<https://www.ncbi.nlm.nih.gov/structure/cdd/wrpsb.cgi>) [24], Pfam (<https://pfam.xfam.org>) [25], and InterProScan (<https://www.ebi.ac.uk/interpro/search/sequence>) [26]. CD Search detects conserved domains within a protein sequence by comparing the query sequence using RPS-BLAST (Reverse Position-Specific BLAST) against position-specific score matrices derived from conserved domain alignments in the Conserved Domain Database (CDD) [27]. Pfam, a protein family database, provides annotations and multiple sequence alignments generated through hidden Markov models (HMMs) [25].

2.4. Multiple Sequence Alignment and Phylogenetic Analysis

A search for protein homologs was conducted using BLASTp from NCBI (<http://www.ncbi.nlm.nih.gov>) against the nonredundant database, employing default parameters. Sequence alignment and phylogenetic tree construction were carried out using the MEGA 11 program [28]. Specifically, the ClustalW algorithm and Maximum Likelihood (ML) technique within MEGA 11 were employed for iterative Multiple Sequence Alignment (MSA) and tree-building processes, respectively.

2.5. Protein Structure Preparation

The secondary structure of the protein was predicted using the PSI-blast based secondary structure prediction (PSIPRED) (<http://bioinf.cs.ucl.ac.uk/psipred>)

[29] and Self-Optimized Prediction Method with Alignment (SOPMA) (https://npsaprabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) [30] servers. The 3D structure of the target protein was determined using the SWISS-MODEL (<https://swissmodel.expasy.org/>) server [31]. This server automatically searches BLASTp to identify suitable templates for each protein sequence. The resulting 3D model structure was visualized using BIOVIA Discovery Studio Visualizer (BIOVIA Discovery Studio 2021). The three-dimensional model structure generated by the SWISS-MODEL server was further refined using the software Swiss-PdbViewer [32].

2.6. Protein Quality Assessment

The quality of the generated model structure was assessed using various evaluation tools, including PROCHECK (<https://www.ebi.ac.uk/thornton-srv/software/PROCHECK>) [33], QMEAN (<https://swissmodel.expasy.org/qmean>) [34] from the ExPASy server of the SWISS-MODEL workspace, and ERRAT (<https://saves.mbi.ucla.edu/>) [35]. Z-scores for both proteins were estimated using the ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>) server [36].

2.7. Protein Active Site Prediction

The Computed Atlas of Surface Topography of Proteins (CASTp) server (<http://sts.bioe.uic.edu/castp/calculation.html>) was employed to identify the predictive protein's active site. It is essential for predicting the regions and critical residues involved in protein-ligand interactions. The CASTp results were visualized using BIOVIA Discovery Studio Visualizer software.

2.8. Subcellular Localization of Protein

The CELLO: Subcellular Localization Predictive System (<http://cello.life.nctu.edu.tw>) [37], Predicts Subcellular Localization of Prokaryotic Proteins (PSLpred) (<https://webs.iitd.edu.in/raghava/pslpred/>) [38], PSORTb v3.0.2 (<https://www.psort.org/psortb/>) [39] and SOSUI (<http://harrier.nagahama-i-bio.ac.jp/sosui>) [40] servers were utilized to predict the subcellular location of the hypothetical protein.

2.9. Molecular Docking Analysis

Docking analysis was conducted using Autodock Vina software (<http://vina.scripps.edu/download.html>) [41], which aids in studying and predicting ligand interactions with macromolecules. The ligand utilized for docking was Arginine beta-naphthylamide which is an inhibitor of ToIC family proteins. Autodock Vina determined the binding affinity between the target protein and ligand [42]. Protein-protein docking between the target protein and the hemolysin-coregulated protein1 (Hcp1) of *S. Typhimurium* was performed using the ClusPro 2.0 server [43]. The docking results were analyzed with Discovery stu-

dio visualizer.

3. Results and Discussions

3.1. Protein Sequence Retrieval

The hypothetical protein identified under the accession number KXA20922.1 originates from the *F. nucleatum* strain MJR7757B. This protein consists of 438 amino acid residues, and its primary sequence was obtained in FASTA format to enable subsequent analysis (Table 1).

3.2. Protein Physicochemical Properties

The putative protein consisted of 438 amino acids and had a molecular weight of 52120.02 Da. It is believed that these amino acids have a half-life of more than 10 hours in bacteria. The pH of the protein is 8.33, indicating a slightly alkaline nature. Their aliphatic index (AI) of 85.89 suggests the presence of aliphatic side chains. The hydrophobicity (GRAVY) has a grand average of -0.823 , showing an average hydrophilic nature. The instability index (II) is 29.47, suggesting a considerable level of stability (Table 2).

3.3. Protein Functional Prediction

Domain analysis involves identifying, characterizing, and understanding the roles of individual domains to gain insights into the overall function and organization of proteins. Several annotation techniques were used to identify conserved regions (domains) and predict the functions of the HP protein. According to the NCBI-CD Search, InterProScan, and Pfam databases the target protein belongs to the outer membrane efflux protein (ToIC family). The ToIC superfamily domain, predicted by the NCBI-CDD server, has an E-value of $7.71e-09$

Table 1. The properties of hypothetical protein retrieved from NCBI database.

Properties	Hypothetical Protein
Locus	KXA20922
Definition	Hypothetical protein HMPREF3221_01179 [<i>F. nucleatum</i>]
Accession	KXA20922
Version	KXA20922.1
Amino acid	438
Organisms	<i>F. nucleatum</i>
FASTA sequence	<p>>KXA20922.1 hypothetical protein HMPREF3221_01179 [<i>F. nucleatum</i>]</p> <p>MIRERMNMKILLFFLILTSLNCSAQETLSIDEALNRVGNDRSEYEFKKFQNSQEGTNV KIKDNKLGDFN GVTLSGGYNISENNFDNRPRKYDRTFQNKATYGPFFVNYNYVQSDRSYVSFGVEKNLKDVFYKYNLNK INNLELNLKISYDKNIQTKKINLVSLYQDILNTKNELEYRKKAYEHYRVLDLKLKSYELGASPKINLE SVELEAEDSKLQIDILETKLKS LYDIGKTDYNIDFENYKLLDFVENNESIDFILNSYMKDEVEELRLSLS MAEERKSYSNYDRYMPDLYLGYERVDRNLRGDRYRQDLFTIKFSKLFSTDSEYKLNLEVENLNKNDL NEKIRVINA EKIKLKSEYHELLKLT SIGDKKSNIAYKKYLIKEKEYELNKSSYLDVIDEYNKYLSQEIET KKAKNALNAFVYKIKIKR</p>

Table 2. The physicochemical properties of hypothetical protein HMPREF3221_01179.

ProtParam Parameters	Values
Number of amino acids	438
Molecular weight (MW)	52120.02 Da
Theoretical pI (Isoelectric point)	8.33
Total number of negatively charged (Asp + Glu)	72
Total number of positively charged (Arg + Lys)	75
Estimated half-life (hr)	>10
Instability index	29.47
Aliphatic index (AI)	85.89
Grade average of hydropathicity (GRAVY)	-0.823
Number of atoms	7357

and is located at amino acid residues 92 - 428. Outer membrane efflux protein (ToIC protein family) has a variety of important functions in bacterial physiology. They aggressively eliminate a range of compounds, such as antibiotics and poisons, serving as a barrier against dangerous chemicals and maintaining cellular homeostasis [44] [45]. Their main documented role is in drug resistance, where they force antibiotics out of cells and so promote multidrug resistance. They engage in interbacterial interactions with certain bacteria (*Escherichia coli*, *Pseudomonas aeruginosa*, and *Salmonella enterica*) by exporting virulence factors or poisons to rival bacteria. They may also contribute to biofilms' production and increase pathogenicity by exporting toxins. Certain efflux proteins move quorum-sensing signalling chemicals [46].

3.4. Sequence Alignment Assessment and Phylogenetic Analysis

According to the NCBI BLASTp search of the target protein in compared to the nonredundant database, the protein shares 98% - 100% sequence similarity with other known ToIC superfamily proteins from different organisms (Table 3). A phylogenetic tree was constructed to depict the relationship between target hypothetical protein and other ToIC family proteins. The BLASTp results were utilized in the construction of the tree by using Mega11 software. The results suggest that most of the proteins are closely related to each other and found a common ancestor (Figure 1).

3.5. Protein Structure Analysis

The results obtained from the SOPMA analysis revealed three conformational states: extended strand (11.64%), alpha helix (60.05%), and random coil (25.11%). The results obtained using PSIPRED showed that the random coil accounted for 25.38% of the structure, the alpha helix accounted for 60%, and the extended strand accounted for 11.77%. The PSIPRED utilized for the prediction of the secondary structure of the protein is shown in Figure 2.

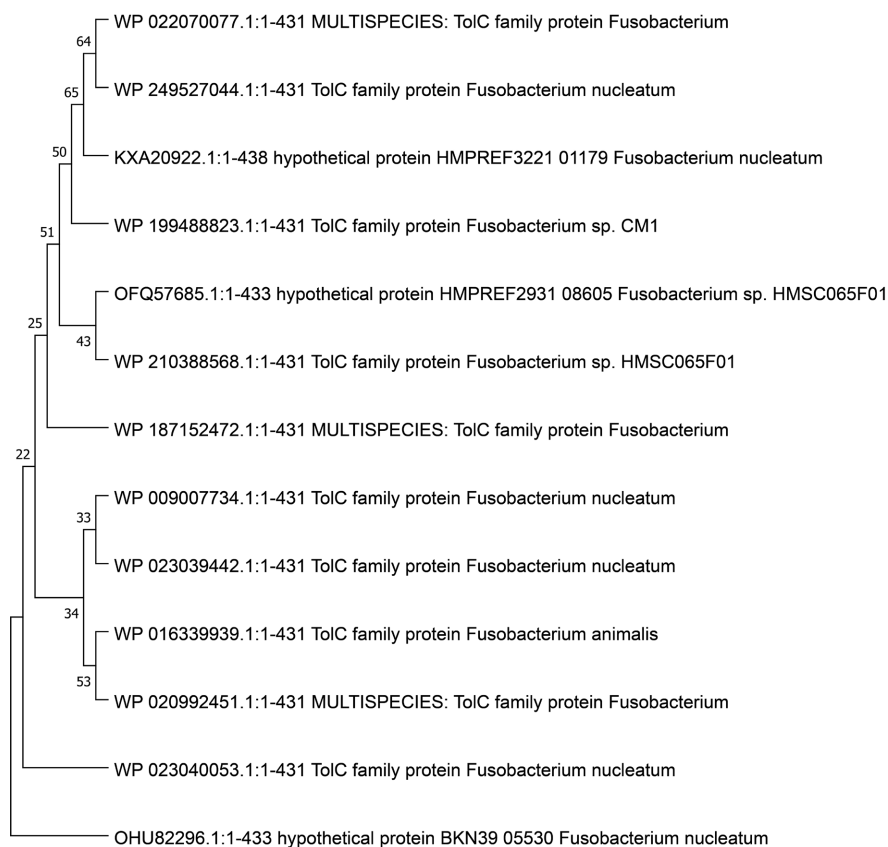


Figure 1. A phylogenetic tree illustrating the target protein's evolutionary relationships with other ToIC proteins. The neighbor-joining approach was used by Mega 11 to create the tree based on the score matrix.

Table 3. NCBI BLASTp result shows thesequence similarity with the target hypotheticalprotein sequence.

Accession	Organism Name	Protein Name	Scores	Per. Identity	E-value
KXA20922.1	<i>Fusobacterium nucleatum</i>	hypothetical protein HMPREF3221_01179	855	100	0
OFQ57685.1	<i>Fusobacterium sp.</i> HMSC065F01	hypothetical protein HMPREF2931_08605	840	99.54	0
WP_022070077.1	<i>Fusobacterium</i>	ToIC family protein	839	100	0
ALF18214.1	<i>Fusobacterium animalis</i>	hypothetical protein RN98_08525	839	99.08	0
WP_249527044.1	<i>Fusobacterium nucleatum</i>	ToIC family protein	838	99.77	0
WP_199488823.1	<i>Fusobacterium sp.</i> CM1	ToIC family protein	838	99.77	0
WP_023040053.1	<i>Fusobacterium nucleatum</i>	ToIC family protein	837	99.3	0
ALF21854.1	<i>Fusobacterium animalis</i>	hypothetical protein RO08_05905	836	98.61	0
WP_210388568.1	<i>Fusobacterium sp.</i> HMSC065F01	ToIC family protein	836	99.54	0
WP_187152472.1	<i>Fusobacterium</i>	ToIC family protein	835	99.3	0

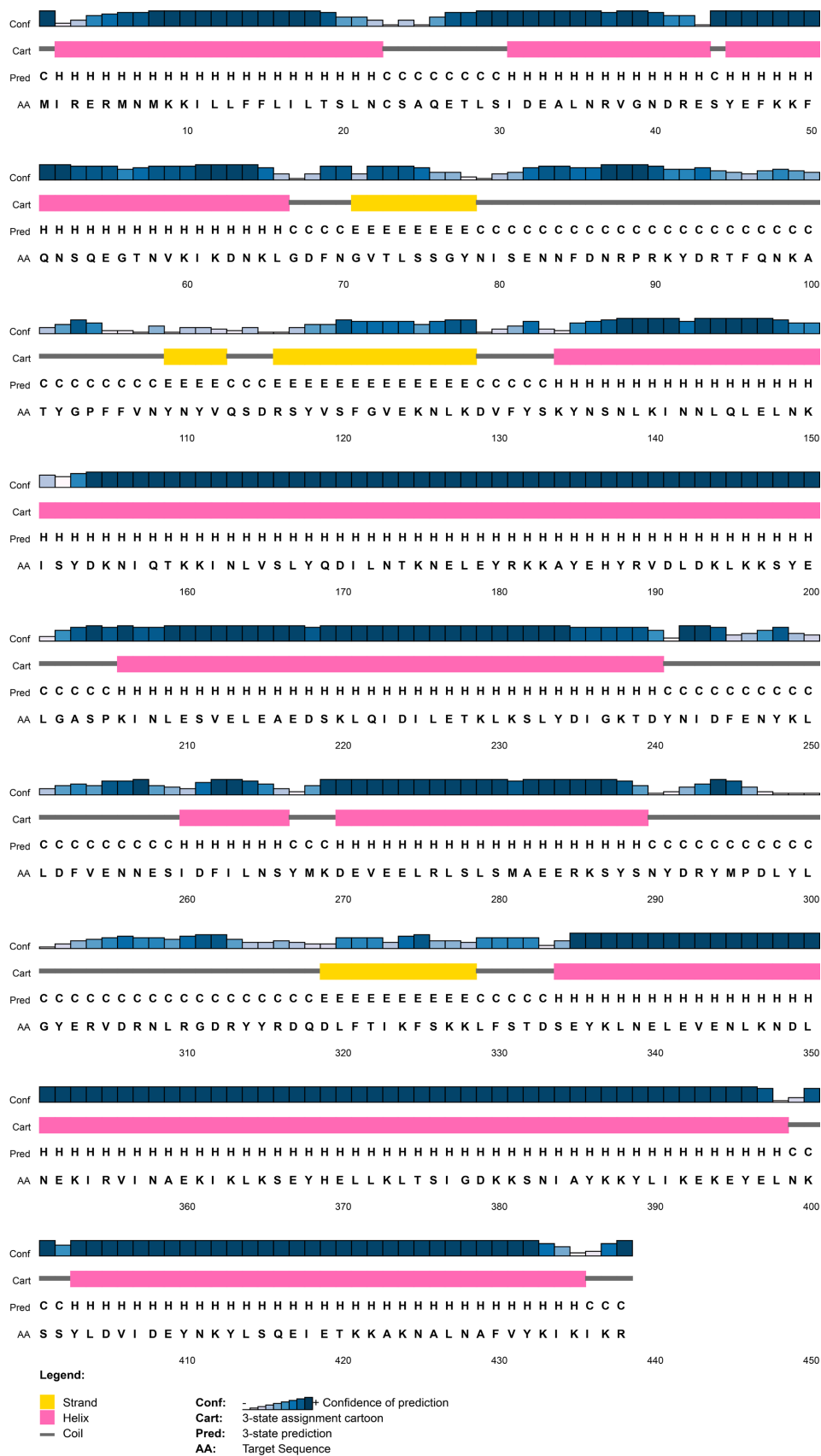


Figure 2. The predicted secondary structure of the target protein by using the PSI-PRED server.

The tertiary structure of the target protein was prepared through SWISS-MODEL service by utilizing a template demonstrating a sequence identity of 93.10% with the hypothetical protein. The Swiss-PdbViewer energy minimization server was utilized for the model protein structure's energy reduction. The 3D structure after energy minimization is shown in Discover studio visualizer (**Figure 3**).

3.6. Quality Assessment of Predicted Structure

Utilizing the SWISS-MODEL service, the protein's three-dimensional (3D) structure was obtained, and it passed all model quality evaluation tools, such as PROCHECK, QMEAN, and ERRAT. As per the PROCHECK results, the ideal area in the Ramachandran plot included 96.6% of the amino acid residues (**Table 4**) (**Figure 4**). The overall residues with a QMEAN4 score of 0.54, regarded as satisfactory (**Figure 5**). Additionally, ERRAT projected that the protein structure had a quality factor of 97.6923, indicating high quality.

The Z-score obtained from the ProSA server showed the model's overall quality. It indicated whether the input structure fell within the range of scores normally found for novel proteins of similar size. The Z score for the model obtained from ProSA was -5.89 (**Figure 6**).

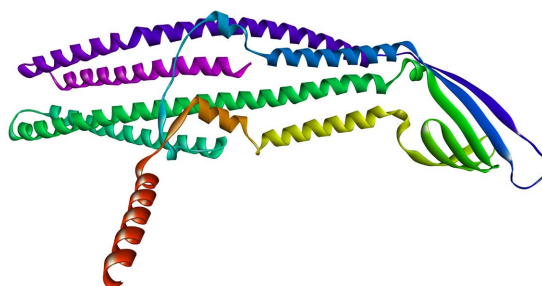


Figure 3. Three-dimensional target protein structure through SWISS-MODEL server after Swiss-PdbViewer energy minimization (visualized by BIOVIA Discovery Studio Visualizer 2021).

Table 4. Ramachandran plots calculations of the target protein.

Plot statistics	Number of AA	Percentage (%)
Residues in most favored regions [A, B, L]	403	96.6
Residues in additional allowed regions [a, b, l, p]	14	3.4
Residues in generously allowed regions [\sim a, \sim b, \sim l, \sim p]	0	0.0
Residues in disallowed regions	0	0.0
Number of non-glycine and non-proline residues	417	100.00
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	12	
Number of proline residues	4	
Total number of residues	435	

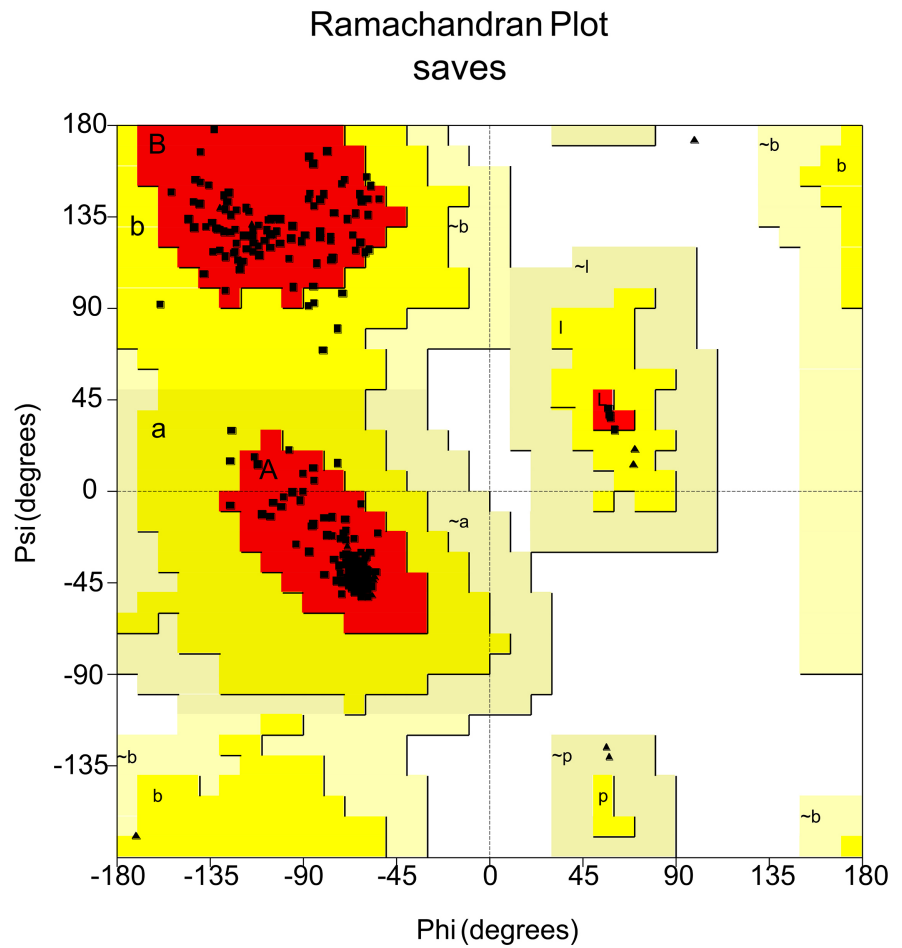


Figure 4. The Ramachandran plot of the model structure, as verified through PROCHECK.

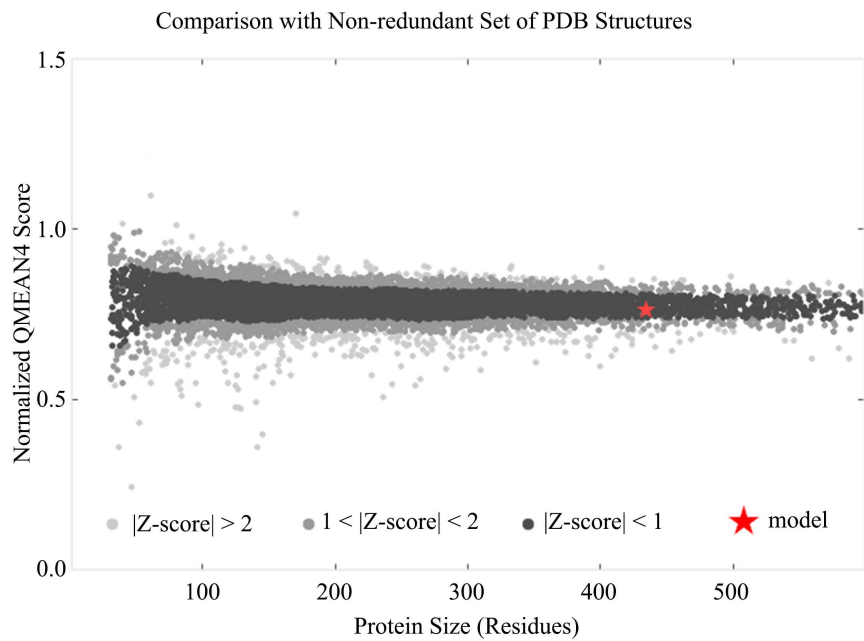


Figure 5. The QMEAN result for the model structure.

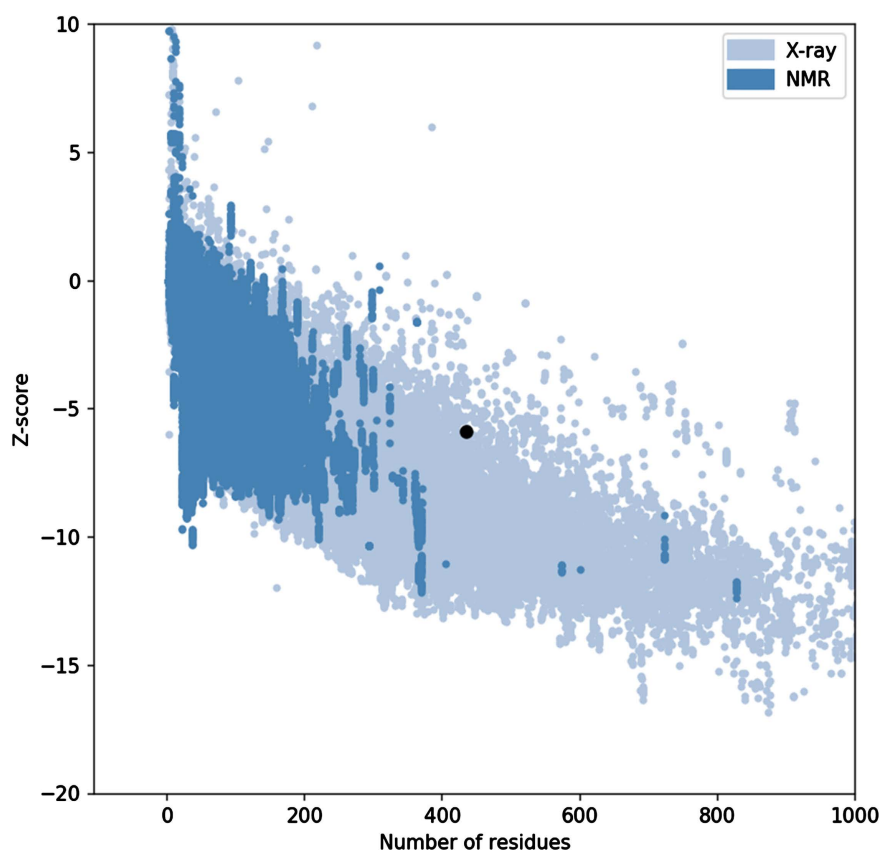


Figure 6. The Z-scores obtained from the ProSA server. The Z-score for the model was -5.89 .

3.7. Active Site Detection

CASTp provides a detailed, comprehensive, and quantitative analysis of a protein's topographical features. It can precisely locate and measure functional pockets on protein surfaces and within the 3D structure's interior. Using the CASTp server, the active site of model structures was examined, and its amino acid residues were ascertained. Then Discover studio was utilized to visualize the results. The major pocket regions were found between 32 - 36, 389 - 396, and 432 - 438, respectively. The model protein's active residues predicted by CASTp are ASP³², LEU³⁵, ASN³⁶, ASP¹⁵⁴, ILE¹⁵⁷, GLN¹⁵⁸, LYS¹⁶¹, ASP²⁷⁰, TYR⁴³², LYS⁴³⁵, ILE⁴³⁶, ARG⁴³⁸ (Figure 7).

3.8. Subcellular Localization of Hypothetical Protein

The CELLO program identified the location of the target protein at outer membrane with a 3.417 reliability score. The findings from PSORTb and PSLpred were also outer membrane and extracellular protein. A putative protein's subcellular location is important since it indicates the function and role that the protein plays within a cell. It provides information on the protein's regulation, interactions with other molecules, and possible role in illness. This knowledge is essential for basic research as well as the creation of new therapeutics [38].

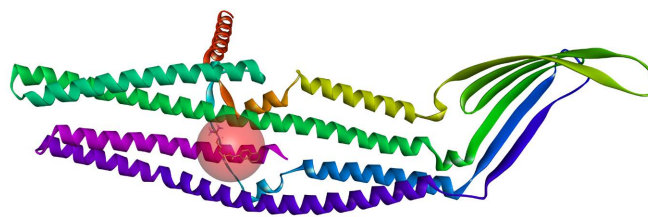


Figure 7. Active site prediction by CASTp server (visualized by BIOVIA Discovery Studio Visualizer 2021).

3.9. Molecular Docking Analysis

Autodock Vina program was utilized to run a docking study between the ligand and the target protein, and the interaction was visualized by Discovery Studio (**Figure 8**). The hypothetical protein belongs to the ToIC protein family which are the efflux proteins that help in pumping the materials across the cell membrane. The compound Arginine beta-naphthylamide is known as an inhibitor of efflux proteins. Therefore, it is employed as a ligand in this work. The ligand demonstrated a substantial affinity for binding to the target hypothetical protein. The ligand's binding affinity for the model was -7.1 kcal/mol (**Table 5**). It was discovered that several of the interaction residues in the proteins' active sites were identical, as predicted by the CASTp server. The discovery of a significant binding affinity of the ligand with the protein of interest further supported our results.

Then the protein-protein interaction of the Hemolysis-coregulated protein 1 (Hcp1) protein of *S. Typhimurium* and the target protein was done by using Cluspro2.0. Hcp1 played an important role in the proper delivery of antibacterial toxins by interacting with efflux proteins. Hence, Hcp1 was utilized in protein-protein interactions. The docking outcomes are mentioned in (**Table 6**). It is noted that maximum residues have taken part in exchange from both proteins. The reason might be the selection of higher cluster members protein-ligand complex from the Cluspro 2.0 server. Experimental research has not yet revealed the precise nature of the interaction between the hcp1 and ToIC proteins. Belonging to the ToIC protein family, renowned for its efflux functions, the protein's interactions with Hcp1 underscore its crucial involvement in the precise delivery of antibacterial toxins.

Overall, the retrieved target protein conserved sequence similar with many *F. nucleatum* species, which supports the efflux protein's potential usage as a therapeutic target. The outer membrane efflux proteins are essential for bacterial major functions. In recent years, progress in understanding these proteins has been increased. To the best of our knowledge, this is the first investigation to describe the structural and functional properties of *F. nucleatum* efflux protein HMPREF3221_01179. We believe this research helps in understating the mechanism of bacterial functions and might help design new drugs in the future. However, more studies are needed to confirm its function at the experimental level.

Table 5. Details of protein-ligand docking analysis.

Protein	Ligand	Binding Affinity (kcal/mol)	Category	Type of Interaction	Key Interacting Residues
Hypothetical protein HMPREF3221_01179	Arginine beta-naphthylamide	-7.1	Hydrogen Bond	Conventional H bond, Pi-alkyl	Ile162, Ser166, Gln169, Asp170, Glu367, Lys433, Lys437

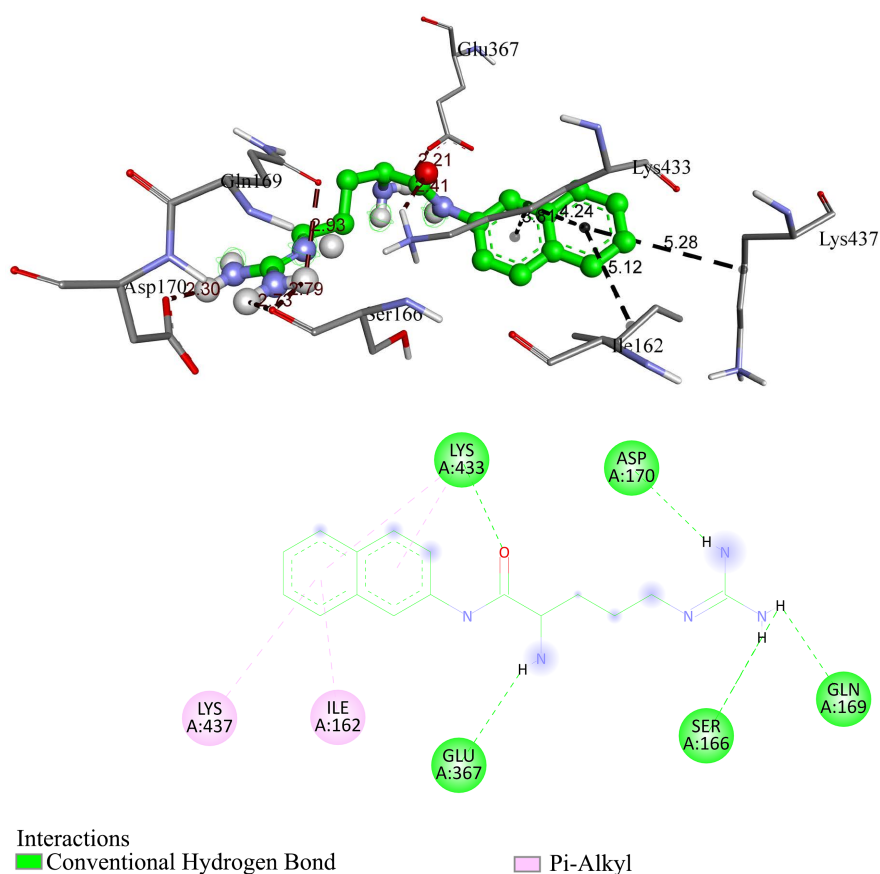


Figure 8. The 3D and 3D interactions of hypothetical protein HMPREF3221_01179 and ligand meso-diaminopimelic acid. The green colour stick represents the Arginine beta-naphthylamide ligand, which has been docked in the active region of the target protein (Top). The bonding between amino acids and the ligand (Bottom). The docking is visualized by Discovery studio visualizer).

Table 6. Protein-protein interaction analysis.

Receptor	Ligand	Cluster Members	Weighted Energy Score of The Centre
Hypothetical protein KXA20922.1	Hypothetical protein PA0085	75	-1071

4. Conclusion

Microbial genome hypothetical proteins study is crucial for unravelling their

unknown functions, leading to insights into microbial biology, potential drug targets, and applications in biotechnology. This in-depth analysis of the hypothetical protein HMPREF3221_01179 from *F. nucleatum* strain MJR7757B provides valuable insights into its structural, functional, and interaction properties, suggesting its potential as a therapeutic target. Additionally, these findings unveil opportunities for further exploration of this bacterium in the realm of biotechnological applications.

Authors Contribution

Conceptualization: Md. Isrifil Hossen, Sayed Mashequl Bari. Methodology: Sayed Mashequl Bari, Md. Isrifil Hossen, Nusrat Jahan. Formal analysis: Sayed Mashequl Bari, Md. Isrifil Hossen, Nusrat Jahan, Fouzia Mostafa. Writing original draft: Sayed Mashequl Bari, Md. Isrifil Hossen, Fouzia Mostafa. Writing review & editing: Amgad Albahi, Jannatul Ferdous.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

References

- [1] Shang, F.-M. and Liu, H.-L. (2018) *Fusobacterium nucleatum* and Colorectal Cancer: A Review. *WJGO*, **10**, 71-81. <https://doi.org/10.4251/wjgo.v10.i3.71>
- [2] Alon-Maimon, T., Mandelboim, O. and Bachrach, G. (2022) *Fusobacterium nucleatum* and Cancer. *Periodontology*, **89**, 166-180. <https://doi.org/10.1111/prd.12426>
- [3] Chen, Y., Shi, T., Li, Y., Huang, L. and Yin, D. (2022) *Fusobacterium nucleatum*: The Opportunistic Pathogen of Periodontal and Peri-Implant Diseases. *Frontiers in Microbiology*, **13**, Article 860149. <https://doi.org/10.3389/fmicb.2022.860149>
- [4] Han, Y.W. (2015) *Fusobacterium nucleatum*: A Commensal-Turned Pathogen. *Current Opinion in Microbiology*, **23**, 141-147. <https://doi.org/10.1016/j.mib.2014.11.013>
- [5] Allen-Vercoe, E., Strauss, J. and Chadee, K. (2011) *Fusobacterium nucleatum*: An Emerging Gut Pathogen? *Gut Microbes*, **2**, 294-298. <https://doi.org/10.4161/gmic.2.5.18603>
- [6] Bashir, A., Miskeen, A.Y., Bhat, A., Fazili, K.M. and Ganai, B.A. (2015) *Fusobacterium nucleatum*: An Emerging Bug in Colorectal Tumorigenesis. *European Journal of Cancer Prevention*, **24**, 373-385. <https://doi.org/10.1097/CEJ.0000000000000116>
- [7] Storm, J.C., Ford, B.A. and Streit, J.A. (2013) Myocardial Infection Due to *Fusobacterium nucleatum*. *Diagnostic Microbiology and Infectious Disease*, **77**, 373-375. <https://doi.org/10.1016/j.diagmicrobio.2013.08.022>
- [8] Nwaokorie, F.O., Coker, A.O., Ogunsola, F.T., Avika-Campos, M.J., Gaetti-Jardim, E., Ayanbadejo, P.O., Umeizudike, K.A. and Abdurrazaq, O.T. (2011) Isolation and Molecular Identification of *Fusobacterium nucleatum* from Nigerian Patients with Oro-Facial Infections. *West African Journal of Medicine*, **30**, 125-129.
- [9] Le Monnier, A., Jamet, A., Carbonnelle, E., Barthod, G., Moumile, K., Lesage, F., Zahar, J.-R., Mannach, Y., Berche, P. and Couloigner, V. (2008) *Fusobacterium Neocrophorum* Middle Ear Infections in Children and Related Complications: Report of 25 Cases and Literature Review. *The Pediatric Infectious Disease Journal*, **27**,

- 613-617. <https://doi.org/10.1097/INF.0b013e318169035e>
- [10] Stergiopoulou, T. and Walsh, T.J. (2016) *Fusobacterium necrophorum* Otitis and Mastoiditis in Infants and Young Toddlers. *European Journal of Clinical Microbiology & Infectious Diseases*, **35**, 735-740. <https://doi.org/10.1007/s10096-016-2612-1>
- [11] Naveed, M., Makhdoom, S.I., Abbas, G., Safdari, M., Farhadi, A., Habtemariam, S., Shabbir, M.A., Jabeen, K., Asif, M.F. and Tehreem, S. (2022) The Virulent Hypothetical Proteins: The Potential Drug Target Involved in Bacterial Pathogenesis. *Mini-Reviews in Medicinal Chemistry*, **22**, 2608-2623. <https://doi.org/10.2174/1389557522666220413102107>
- [12] Zhao, J., Cao, Y. and Zhang, L. (2020) Exploring the Computational Methods for Protein-Ligand Binding Site Prediction. *Computational and Structural Biotechnology Journal*, **18**, 417-426. <https://doi.org/10.1016/j.csbj.2020.02.008>
- [13] Dukka, B.K. (2013) Structure-Based Methods for Computational Protein Functional Site Prediction. *Computational and Structural Biotechnology Journal*, **8**, e201308005. <https://doi.org/10.5936/csbj.201308005>
- [14] Mills, C.L., Beuning, P.J. and Ondrechen, M.J. (2015) Biochemical Functional Predictions for Protein Structures of Unknown or Uncertain Function. *Computational and Structural Biotechnology Journal*, **13**, 182-191. <https://doi.org/10.1016/j.csbj.2015.02.003>
- [15] Watson, J.D., Laskowski, R.A. and Thornton, J.M. (2005) Predicting Protein Function from Sequence and Structural Data. *Current Opinion in Structural Biology*, **15**, 275-284. <https://doi.org/10.1016/j.sbi.2005.04.003>
- [16] Valencia, A. (2005) Automatic Annotation of Protein Function. *Current Opinion in Structural Biology*, **15**, 267-274. <https://doi.org/10.1016/j.sbi.2005.05.010>
- [17] Espadaler, J., Querol, E., Aviles, F.X. and Oliva, B. (2006) Identification of Function-Associated Loop Motifs and Application to Protein Function Prediction. *Bioinformatics*, **22**, 2237-2243. <https://doi.org/10.1093/bioinformatics/btl382>
- [18] Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A. and Wheeler, D.L. (2002) GenBank. *Nucleic Acids Research*, **30**, 17-20. <https://doi.org/10.1093/nar/30.1.17>
- [19] The UniProt Consortium (2023) UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, **51**, D523-D531.
- [20] Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. (2003) ExPASy: The Proteomics Server for In-Depth Protein Knowledge and Analysis. *Nucleic Acids Research*, **31**, 3784-3788. <https://doi.org/10.1093/nar/gkg563>
- [21] Kyte, J. and Doolittle, R.F. (1982) A Simple Method for Displaying the Hydrophobic Character of a Protein. *Journal of Molecular Biology*, **157**, 105-132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
- [22] Gill, S.C. and von Hippel, P.H. (1989) Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. *Analytical Biochemistry*, **182**, 319-326. [https://doi.org/10.1016/0003-2697\(89\)90602-7](https://doi.org/10.1016/0003-2697(89)90602-7)
- [23] Henriksson, G., Englund, A.K., Johansson, G. and Lundahl, P. (1995) Calculation of the Isoelectric Points of Native Proteins with Spreading of pK_a Values. *Electrophoresis*, **16**, 1377-1380. <https://doi.org/10.1002/elps.11501601227>
- [24] Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., *et al.* (2017) CDD/SPARCLE: Functional Classification of Proteins via Subfamily Domain Architectures. *Nucleic Acids Re-*

- search*, **45**, D200-D203. <https://doi.org/10.1093/nar/gkw1129>
- [25] 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., *et al.* (2021) Pfam: The Protein Families Database in 2021. *Nucleic Acids Research*, **49**, D412-D419. <https://doi.org/10.1093/nar/gkaa913>
- [26] Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R. (2005) InterProScan: Protein Domains Identifier. *Nucleic Acids Research*, **33**, W116-W120. <https://doi.org/10.1093/nar/gki442>
- [27] Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., *et al.* (2015) CDD: NCBI's Conserved Domain Database. *Nucleic Acids Research*, **43**, D222-D226. <https://doi.org/10.1093/nar/gku1221>
- [28] Kumar, S., Nei, M., Dudley, J. and Tamura, K. (2008) MEGA: A Biologist-Centric Software for Evolutionary Analysis of DNA and Protein Sequences. *Briefings in Bioinformatics*, **9**, 299-306. <https://doi.org/10.1093/bib/bbn017>
- [29] Buchan, D.W.A. and Jones, D.T. (2019) The PSIPRED Protein Analysis Workbench: 20 Years On. *Nucleic Acids Research*, **47**, W402-W407. <https://doi.org/10.1093/nar/gkz297>
- [30] Combet, C., Blanchet, C., Geourjon, C. and Deléage, G. (2000) NPS@: Network Protein Sequence Analysis. *Trends in Biochemical Sciences*, **25**, 147-150. [https://doi.org/10.1016/S0968-0004\(99\)01540-6](https://doi.org/10.1016/S0968-0004(99)01540-6)
- [31] Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., *et al.* (2018) SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Research*, **46**, W296-W303. <https://doi.org/10.1093/nar/gky427>
- [32] Kaplan, W. and Littlejohn, T.G. (2001) Swiss-PDB Viewer (Deep View). *Briefings in Bioinformatics*, **2**, 195-197. <https://doi.org/10.1093/bib/2.2.195>
- [33] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography*, **26**, 283-291. <https://doi.org/10.1107/S0021889892009944>
- [34] Benkert, P., Biasini, M., and Schwede, T. (2011) Toward the Estimation of the Absolute Quality of Individual Protein Structure Models. *Bioinformatics*, **27**, 343-350. <https://doi.org/10.1093/bioinformatics/btq662>
- [35] Colovos, C. and Yeates, T.O. (1993) Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Science*, **2**, 1511-1519. <https://doi.org/10.1002/pro.5560020916>
- [36] Wiederstein, M. and Sippl, M.J. (2007) ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins. *Nucleic Acids Research*, **35**, W407-W410. <https://doi.org/10.1093/nar/gkm290>
- [37] Yu, C.-S., Chen, Y.-C., Lu, C.-H. and Hwang, J.-K. (2006) Prediction of Protein Subcellular Localization. *Proteins*, **64**, 643-651. <https://doi.org/10.1002/prot.21018>
- [38] Bhasin, M., Garg, A. and Raghava, G.P.S. (2005) PSLpred: Prediction of Subcellular Localization of Bacterial Proteins. *Bioinformatics*, **21**, 2522-2524.
- [39] Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J., *et al.* (2010) PSORTb 3.0: Improved Protein Subcellular Localization Prediction with Refined Localization Subcategories and Predictive Capabilities for All Prokaryotes. *Bioinformatics*, **26**, 1608-1615.

- <https://doi.org/10.1093/bioinformatics/btq249>
- [40] Imai, K., Asakawa, N., Tsuji, T., Akazawa, F., Ino, A., Sonoyama, M. and Mitaku, S. (2008) SOSUI-GramN: High Performance Prediction for Sub-Cellular Localization of Proteins in Gram-Negative Bacteria. *Bioinformatics*, **2**, 417-421. <https://doi.org/10.6026/97320630002417>
- [41] Trott, O. and Olson, A.J. (2010) AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *Journal of Computational Chemistry*, **31**, 455-461. <https://doi.org/10.1002/jcc.21334>
- [42] Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N. and Alva, V. (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at Its Core. *Journal of Molecular Biology*, **430**, 2237-2243. <https://doi.org/10.1016/j.jmb.2017.12.007>
- [43] Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D. and Vajda, S. (2017) The ClusPro Web Server for Protein-Protein Docking. *Nature Protocols*, **12**, 255-278. <https://doi.org/10.1038/nprot.2016.169>
- [44] Zgurskaya, H.I., Krishnamoorthy, G., Ntrel, A. and Lu, S. (2011) Mechanism and Function of the Outer Membrane Channel TolC in Multidrug Resistance and Physiology of Enterobacteria. *Frontiers in Microbiology*, **2**, Article 189. <https://doi.org/10.3389/fmicb.2011.00189>
- [45] Masi, M. and Pagès, J.-M. (2013) Structure, Function and Regulation of Outer Membrane Proteins Involved in Drug Transport in Enterobacteriaceae: The OmpF/C-TolC Case. *TOMICROJ*, **7**, 22-33. <https://doi.org/10.2174/1874285801307010022>
- [46] Kumar, S. and Varela, M.F. (2012) Biochemistry of Bacterial Multidrug Efflux Pumps. *International Journal of Molecular Sciences*, **13**, 4484-4495. <https://doi.org/10.3390/ijms13044484>