

Structural and Functional Characterization of a Putative Type VI Secretion System Protein in *Cronobacter sakazakii* as a Potential Therapeutic Target: A Computational Study

Evolutionary Bioinformatics
Volume 21: 1–17
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11769343251327660



Nurun Nahar Akter¹ , Md. Moin Uddin² , Nesar Uddin³,
Israt Jahan Asha¹ , Md Soyeb Uddin⁴, Md. Arju Hossain⁵,
Fahadul Alam⁶, Siratul Kubra Shifat⁷, Md. Abu Zihad⁸
and Md Habibur Rahman^{9,10}

¹Department of Biotechnology and Genetic Engineering, Noakhali Science and Technology University, Noakhali, Bangladesh. ²Department of Biotechnology, Bangladesh Agricultural University, Mymensingh, Bangladesh. ³Department of Apparel Engineering, Textile Engineering College, Noakhali, Bangladesh. ⁴Department of Botany, Cox's Bazar Government College, Cox's Bazar, Bangladesh. ⁵Department of Biochemistry and Biotechnology, Khwaja Yunus Ali University, Sirajganj, Bangladesh. ⁶Department of Biological Sciences, Texas Tech University, TX, USA. ⁷Department of Bioprocess and Genetic Engineering, Chattogram Veterinary and Animal Sciences University, Chattogram, Bangladesh. ⁸Department of Microbiology, Primeasia University, Dhaka, Bangladesh. ⁹Department of Computer Science and Engineering, Islamic University, Kushtia, Bangladesh. ¹⁰Center for Advanced Bioinformatics and Artificial Intelligence Research, Islamic University, Kushtia, Bangladesh.

ABSTRACT

BACKGROUND: *Cronobacter sakazakii*, a foodborne pathogen with a fatality rate of 33%, is a rod-shaped, Gram-negative, non-spore-forming bacterium responsible for causing meningitis, bacteremia, and necrotizing enterocolitis. Despite many unknown functions of hypothetical proteins in bacterial genomes, bioinformatic techniques have successfully annotated their roles in various pathogens.

OBJECTIVES: The aim of this investigation is to identify and annotate the structural and functional properties of a hypothetical protein (HP) from *Cronobacter sakazakii* 7G strain (accession no. WP_004386962.1, 277 residues) using computational tools.

METHODS: Multiple bioinformatic tools were used to identify the homologous protein and to construct and validate its 3D structure. A 3D model was generated using SWISS-MODEL and validated using tools, developing a reliable 3D structure. The STRING and CASTp servers provided information on protein-protein interactions and active sites, identifying functional partners.

RESULTS: The putative protein was soluble, stable, and localized in the cytoplasmic membranes, indicating its biological activity. Functional annotation identified TagJ (HsiE1) within the protein, a member of the ImpE superfamily involved in the transport of toxins and a part of the bacterial type VI secretion system (T6SS). The 3-dimensional structure of this protein was validated through molecular docking involving 6 different compounds. Among these, ceforanide demonstrated the strongest binding scores, -7.5 kcal/mol for the hypothetical protein and -7.2 kcal/mol for its main template protein (PDB ID: 4UQX.1).

CONCLUSION: Comparative genomics study suggests that the protein found in *C. sakazakii* may be a viable therapeutic target because it seems distinctive and different from human proteins. The results of multiple sequence alignment (MSA) and molecular docking supported HP's potential involvement as a T6SS. These in silico results represent that the examined HP could be valuable for studying *C. sakazakii* infections and creating medicines to treat *C. sakazakii*-mediated disorders.

KEYWORDS: *Cronobacter sakazakii*, hypothetical protein, *in-silico*, T6SS, SWISS-MODEL

RECEIVED: October 3, 2024. **ACCEPTED:** February 28, 2025.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Md. Habibur Rahman, Center for Advanced Bioinformatics and Artificial Intelligence Research, Department of Computer Science and Engineering, Islamic University, Kushtia 7003, Bangladesh. Email: habib@iu.ac.bd

Introduction

Cronobacter sakazakii is a rod-shaped, motile, Gram-negative bacterium belonging to the Enterobacteriaceae family. While it infrequently causes illness in healthy adults, it has been linked

with outbreaks in neonates (particularly premature infants), as well as isolated cases in highly immunocompromised individuals and the elderly.¹ *Cronobacter sakazakii* (*C. sakazakii*) has different kinds of strains, among all these strains *Cronobacter*



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

sakazakii 7G is one of the strains. Its genome size is 4.3 Mb, having total gene number 4122, and its protein-coding gene number is 3899.² It can live in extremely dry conditions found in products like baby formula, protein shakes, powdered milk, and other dried foods.^{3,4} *C. sakazakii* infections are rare, they have been associated with meningitis and sepsis, particularly in newborns and young children.^{5,6} Due to its acid tolerance, *C. sakazakii* exhibits significant resistance to low pH environments. Additionally, this capacity of the bacterium to form bio-film enhances its resistance to antibiotics.⁷ Premature babies may be particularly vulnerable because their stomach acid is not fully developed, even though adult gastric juice normally has a pH between 2 and 3.⁵

Epidemiological studies have linked powdered infant formula (PIF) to over 90% of these *Cronobacter* illnesses. Contamination of PIF can occur during the manufacturing process, recovery, or storage.⁸ Still, some *Cronobacter* species—*C. sakazakii* in particular—benefit from the growing conditions provided by dehydrated formulations. These bacteria are considered to be highly pathogenic, and their presence in baby formula raises serious concerns for public health since they can cause serious infections like meningitis and enterocolitis.^{5,9} Neonates comprise the majority of *C. sakazakii* infection cases. Reports indicate that case fatalities from necrotizing enterocolitis, meningitis, and sepsis can reach up to 33%.¹⁰ *C. sakazakii* will therefore be the focus of this study. Despite the rarity of *Cronobacter* infections, the mortality rate can range from 33% to 80%.^{5,11,12} Newborns who survive *Cronobacter* infection often experience severe long-term effects, such as cerebral abscesses, developmental delays, and loss of hearing and vision.¹³

The majority of disease-causing bacteria have become resistant to several medications.¹⁴ *Cronobacter* species exhibit a higher degree of antibiotic resistance compared to other Enterobacteriaceae members. While limited studies have documented the presence of multi-drug resistance (MDR) in *Cronobacter* isolates from clinical and environmental sources, the underlying molecular mechanisms responsible for their antibiotic resistance remain largely unexplored.¹⁵ It is recommended to treat neonatal infections caused by *C. sakazakii* with antibiotic therapy.¹⁶ Most cases of *Cronobacter* spp. infections respond well to antibiotic therapy¹⁷ and the conventional treatment typically involves a combination of ampicillin and gentamicin or chloramphenicol.³ However, the proliferation of antibiotic-resistant strains has increased due to the overuse and improper administration of medications.¹⁶ To address the issue of multidrug-resistant bacteria, it is crucial to identify new therapeutic targets given the proliferation of antibiotic-resistant strains.¹⁴ Traditional drug development methods are characterized by substantial time and financial commitments.¹⁸ Computer-based analysis utilizes various techniques to point out potential drugs and vaccines, design molecules

based on structure, evaluate efficacy, study host-pathogen interactions, and conduct genome-based comparative research. This approach minimizes the need for extensive laboratory experiments.^{19,20}

Next-generation sequencing (NGS) facilitates rapid data collection, yet determining gene functions remains a significant challenge.^{21,22} A significant proportion of bacterial genomes, ranging from 30% to 40%, consists of genes that are currently classified as hypothetical or of unknown function.²³ The biochemical and functional characterization of these hypothetical proteins is crucial for validating their roles. Bioinformatics plays a crucial role in annotating the functional characteristics of these hypothetical proteins, a process essential for drug discovery and development. Identifying these unknown proteins *in-silico* can improve the effectiveness of drugs and vaccines by uncovering biochemical pathways essential for bacterial survival and pathogenicity.^{24,25} This study utilized different bioinformatics tools to analyze the structural and functional aspects, as well as molecular docking, of a putative protein (accession no. WP_004386962.1) from *C. sakazakii* 7G. Figure 1 illustrates the overall procedure of this study.

Materials and Methods

Retrieval of FASTA (Fast alignment search Tool-All) sequence

About 1906 *C. sakazakii* genomes are accessible in the NCBI (<http://www.ncbi.nlm.nih.gov/>)²⁶ database. For this investigation, we filtered 257 hypothetical proteins from *C. sakazakii* and then we focused on a hypothetical protein composed of 277 amino acids, obtained from the *C. sakazakii* 7G strain (WP_004386962.1). The primary sequence of the protein was retrieved in FASTA format for subsequent analysis. Besides, we clarified that hypothetical proteins were filtered based on their annotation status in the NCBI database, with a focus on proteins of appropriate sequence length and completeness for downstream analysis. Additionally, we explicitly stated that proteins lacking sufficient annotation or with incomplete sequences were excluded in our study.

Exploration of physicochemical characteristics

The physicochemical properties of the protein were evaluated using the ProtParam tool (<http://web.expasy.org/protparam>)²⁷ available on the ExPASy server. The ProtParam tool calculates different physicochemical characteristics of proteins, which provide insights into their functional and structural characteristics. By analyzing a single protein sequence, ProtParam provided insights into the physicochemical properties of a protein, including molecular mass, stability, charge, amino acid composition, and hydrophobicity. This tool offers an accurate and efficient method for assessing the physical and chemical attributes of proteins.²⁸

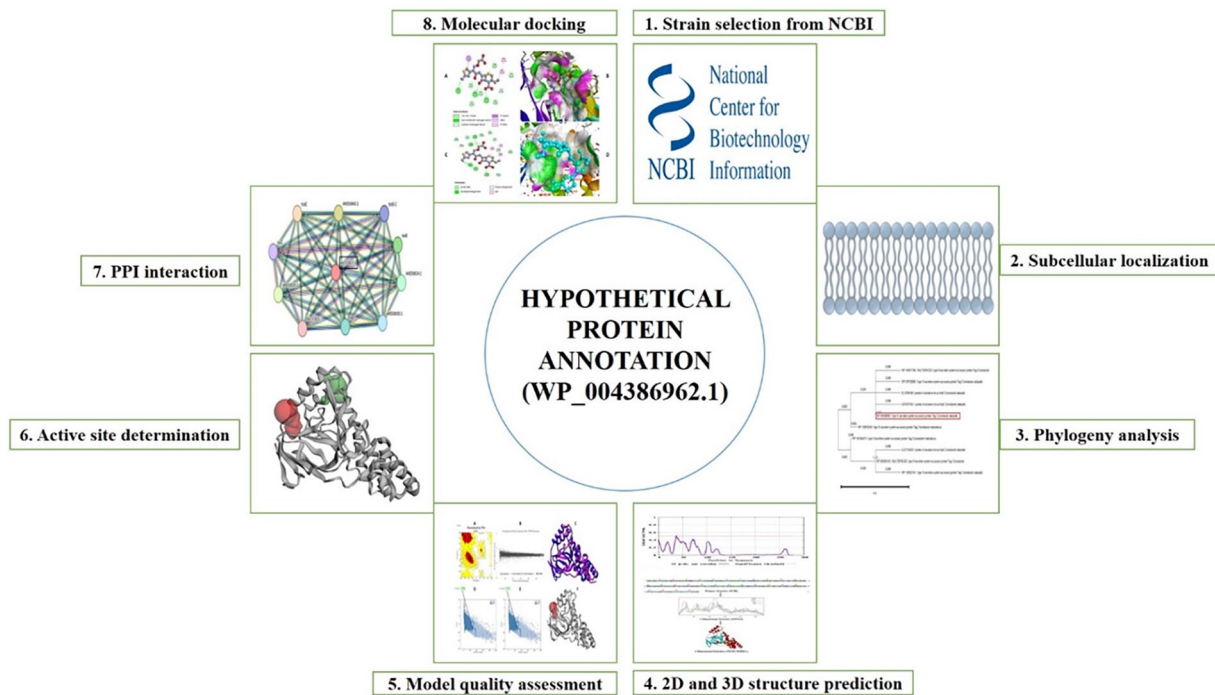


Figure 1. An overview of putative protein annotation and the findings that followed.

Multiple sequence alignment (MSA) and phylogenetic tree evaluation

For the HP homologs (WP_004386962.1), a BLASTp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>; Retrieved date: 5 Jan 2024)²⁹ search was performed against the NCBI non-redundant database using default parameters to identify homologous proteins. For determining homologous proteins, this method finds protein alignments locally. This study used MEGA 11 (version 11.0.13) software for Multiple sequence alignment and phylogenetic tree evaluation.³⁰ Using the FASTA sequence of our query protein (WP_004386962.1), along with 9 related proteins, we performed a multiple sequence alignment, WP_004386962.1, and homologous 9 proteins, WP_105653240.1, WP_063265155.1, WP_104671789.1, WP_076728080.1, WP_161584471.1, WP_105622741.1, ELY2789108.1, EGT4277351.1, and EJC1154322.1. Recent advancements in sequencing alignments have improved precision, scalability, and the ability to compare proteins with different domain designs, making them more efficient for inferring phylogenies and predicting protein structure.³¹

Prognosis of protein solubility and subcellular localization

Understanding the subcellular distribution of proteins is crucial for identifying potential drug or vaccine targets. Membrane proteins can serve as therapeutic or immunological targets, whereas cytoplasmic proteins may be pharmacological targets. CELLO (<http://cello.life.nctu.edu.tw/>)³²

determined the subcellular location of a HP. To further validate the results, we utilized the PSORTb (<https://www.psорт.org/psортb>)³³ and PSLpred (<http://crdd.osdd.net/raghava/pslpred>)³⁴ servers. These servers are commonly employed for forecasting the location of bacterial proteins within cells, providing additional confirmation for our findings. SOSUI (<http://harrier.nagahama-i-bio.ac.jp/sosui/>)³⁵ computes the solubility of a protein and average hydrophobicity.

Determining the motifs and domains of a protein

Domain evaluation was conducted using NCBI Conserved Domain (CD) Search Service (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)³⁶ and InterProScan (<https://www.ebi.ac.uk/interpro/search/sequence/>).³⁷ CD Search can be used to determine which domains in a protein sequence are conserved. RPS-BLAST was utilized to identify conserved domains within the query sequence by comparing it against itself using position-specific matrices from the Conserved Domain Database (CDD).³⁸ The Pfam (<http://pfam.xfam.org/>)³⁹ database, which provides protein family annotations and multiple sequence alignments generated from the Hidden Markov model (HMMs), served as the reference. Protein motifs were identified using the Motif tool on the Genome Net server (<https://www.genome.jp/tools/motif/>).⁴⁰ The coiled-coil structure of the protein was identified using the DeepCoil service.⁴¹ In addition, protein folding patterns were found using the PFP-FunD SeqE server.⁴²

Structure determination and homology modeling of putative protein

The 2-dimensional structure of WP_004386962.1 protein was determined by PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>)⁴³ and SOPMA (https://npsaprabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html).⁴⁴ A correlation was found between the outcome of the PSIPRED study and the SOPMA result. The SWISS-MODEL program was employed to construct a 3D model of the target protein based on homology.⁴⁵ The service utilized BLASTp to identify suitable templates for the entire protein sequence. Based on the search results, the protein 4uqx.1.A was selected as the template for homology modeling. This X-ray diffraction model, with 38.28% sequence identity, represents a possible cytoplasmic protein from *Pseudomonas aeruginosa* and is a good base of reference for modeling. Further protein sequence was evaluated by the HHpred server (<https://toolkit.tuebingen.mpg.de/tools/hhpred>) using MODELLER (version 10.4) where we found 97% sequence identity. The 3D model structure was then displayed with BIOVIA Discovery Studio (version 20.1). A more accurate 3D model was generated using the highest-scoring template. This model was further refined and improved over time using the YASARA energy minimization server using steepest descent method.⁴⁶ It utilizes less energy to produce an accurate and reliable 3-dimensional structure of the target protein.

Quality evaluation of a three-dimensional model

During the last phase of homology modeling, the improved structure of a model protein underwent extensive testing to ensure internal consistency and reliability. Various criteria were utilized to determine the accuracy of this putative protein model. The psi/phi (ψ/ϕ) Ramachandran plot, which was generated by the PROCHECK examination, was used to analyze the backbone configuration. The quality of the model structure was evaluated using the ExpASY service of SWISS-MODEL Workspace's PROCHECK (<https://servicesn.mbi.ucla.edu/PROCHECK/>),⁴⁷ Verify3D (<http://nihserver.mbi.ucla.edu/Verify3D/>),⁴⁸ QMEAN (<https://swissmodel.expasy.org/qmean/>),⁴⁹ and ERRAT (<https://servicesn.mbi.ucla.edu/ERRAT/>)⁵⁰ programs. Moreover, The UCSF Chimera software (version 1.16.0) was used to superimpose and visualize the model and template structure.⁵¹ Also, the ProSA-web server (<https://prosa.services.came.sbg.ac.at/prosa.php>) computed Z scores for both proteins.⁵²

Protein-protein interaction network

Most biological activities are regulated and executed by protein-protein interactions (PPIs), which form a complex, continuous network of reactions. The STRING v11.0 (<https://string-db.org/>)⁵³ inquiry found a protein-protein functional interaction network.

Active site identification

To determine the active site of a protein, the web-based tool CASTp (<http://sts.bioengr.uic.edu/castp>) was used. CASTp is key in identifying, delineating, and quantifying geometric and topological characteristics essential for correcting the protein function. These include interior cavities, cross channels, and surface pockets. Furthermore, it aids in the mapping of functionally annotated residues onto protein 3D structures.⁵⁴

Molecular docking analysis

Preparation of receptor and ligand. The receptor for molecular docking was chosen to be the putative TagJ protein (accession no. WP_004386962.1). Following that, for comparing our binding affinities, the 3D structure of the primary template TagJ protein (PDB ID: 4uqx.1)⁵⁵ was examined using the RCSB PDB server (<http://www.pdb.org>).⁵⁶ After recognizing and removing co-crystallized compounds from the 4uqx.1 structure, within the 3-dimensional coordinate file crystallographic water molecules were retrieved.

In this *in-silico* docking experiment, we employed different anti-bacterial FDA-approved Therapeutics as ligands such as ceforanide (CID:43507), ceftriaxone (CID:5479530), latamoxef (CID: 47499), cefixime (CID: 5362065), pefloxacin (CID: 51081), and amikacin (CID: 37768). The choice of these ligands was guided by a literature review study, focusing on their proven antibacterial effectiveness in favor of a diverse range of Gram-negative bacteria, especially those from the Enterobacteriaceae family.^{57,58} Given that our selected hypothetical protein (WP_004386962.1) from *C. sakazakii* is a Gram-negative bacterium within the Enterobacteriaceae family, these compounds are promising candidates as inhibitors for treating this bacterial infection. The decision to use these inhibitors over other compounds was motivated by our goal to validate the 3D model of the TagJ protein from *C.sakazakii* and compare it with the template TagJ protein from *Pseudomonas aeruginosa*.

After retrieving the canonical SMILES ID of a drug from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov>),⁵⁹ we converted the 3-dimensional SDF to the PDB structure of a drug using Pymol software.⁶⁰ Following this, we significantly simplified the research process by performing ligand optimization and converting it to the PDBQT files utilizing Autodock Vina Tools 1.5.7.⁶¹

Molecular docking and binding interactions analysis. Autodock Vina facilitates the docking of small-molecule libraries to macromolecules, enabling the identification of lead compounds with specific biological functions. The Auto Grid engine generates the configuration file for grid specifications.⁶¹ During the docking process, possible candidates for favored binding were identified using an average root-mean-square deviation (RMSD) of below 1 kcal/mol. The ligands with the highest binding affinity have the highest negative binding energy. For site-specific docking, the grid box was built to the following

Table 1. The ProtParam tool was employed to analyze the physicochemical characteristics of the hypothetical protein (WP_004386962.1).

CHARACTERISTICS	VALUE OF WP_004386962.1 PROTEIN
Number of amino acids	277
Molecular weight	30427.40 Da
Theoretical pI	4.54
Total number of negatively charged residues (Asp + Glu)	37
Total number of positively charged residues (Arg + Lys)	20
Extinction coefficient	280 M ⁻¹ cm ⁻¹
Instability index (II)	31.85
Aliphatic index	92.45
Grand average of hydropathicity (GRAVY)	−0.140

specifications: dimension $x:y:z = 50:50:50$, exhaustiveness = 8, and center $x:y:z = 9.175:-8.927: .75$. Subsequently, hydrophobic and hydrogen bond interactions were analyzed using Biovia Discovery Studio.⁶²

Comparative genomics approach. To find human proteins that might be similar to the hypothetical protein WP_004386962.1, we used a BLASTp search of the human protein database. We employed stringent filtering criteria, setting a minimum bit score of 100 and an *E*-value cutoff of 0.005 to ensure the reliability of the identified hits.⁶³

Results

Physicochemical properties and subcellular localization

To understand the function and molecular evolution of a protein requires a comprehensive assessment of its physicochemical characteristics. With 277 amino acids, this protein exhibits a diverse composition. Among its constituents, Cys (3), Tyr (4), His (4), Lys (4), Met (5), Ile (7), Asn (7), Thr (9), Trp (9), Phe (9), Ser (11), Val (15), Arg (16), Asp (17), Pro (17), Gly (18), Gln (20), Glu (20), Leu (37), and Ala (41) are notably abundant, giving protein negative charge is evident, molecular weight of 30427.40 Da and a putative isoelectric point (pI) of 4.54. Table 1 presents a comprehensive overview of the physicochemical characteristics of the putative protein, encompassing its frequency and predominance with their constituent amino acids. According to the CELLO program, our target protein was predicted to localize to the “Cytoplasmic” region. Subsequent analysis using the PSORTb and PSLpred protein subcellular localization services also yielded strong localization scores, further

confirming its cytoplasmic classification. Additionally, the SOSUI server projected the protein to be soluble.

Phylogenetic tree construction and multiple sequences alignment

In our investigation, we employed BLASTp to survey nonredundant databases, revealing similarities with other bacterial type VI secretion system (T6SS) auxiliary proteins (Table 2). To delve deeper into these similarities, we aligned the FASTA sequence of the putative protein (WP_004386962.1) with the FASTA sequences of related annotated proteins (Supplemental Figure S1). Using several Type VI secretion system accessory protein TagJ protein sequences, we generated a phylogenetic tree using the MEGA11 program. Most of the TagJ proteins utilized in building the phylogenetic tree were sourced from *Cronobacter sakazakii* and other *Cronobacter* species, identified through BLASTp analysis. The resulting phylogenetic tree, with branch distances of 0.004, revealed that WP_104671789.1, WP_076728080.1, ELY2789108.1, and EGT4277351.1 were closely related to our query protein (Accession No. WP_004386962.1; Figure 2). Moreover, The *E*-value quantifies the statistical significance of an alignment, indicating how likely it is to be a random occurrence in a database of the specified size. The *E*-value varies with database size and query length. The alignment is best if the *E*-value is near zero.⁶⁴

Domains and motifs analysis

We used many annotation techniques to identify conserved domains of our target protein and potential functionalities. The target protein, TagJ (HsiE1), is a component of the bacterial T6SS and likely contains an ImpE-like domain (Table 3), as suggested by bioinformatics tools like NCBI-CD Search, Pfam, and InterProScan. The structure of this protein secretion system is similar to that of the bacteriophage puncturing mechanism. This protein is commonly found and contributes to the development of various opportunistic infections. T6SS enhances pathogen survival by injecting protein effectors into host cells and releasing toxins to nearby pathogens.⁶⁵ ClpV and the TssC/TssB sheath interact with it in a particular way.⁵⁵ Bacterial complexes that resemble phage tails are called type VI secretion systems, or T6SS for brief. When the TssBC sheath contracts, the toxin delivery mechanism releases poisons into the target cells.⁶⁶ T6SSs possess 13 essential proteins. These proteins form a puncturing complex similar to that found in bacteriophages. The complex includes a tube, a puncturing tip, and a contractile sheath. Other proteins help assemble the baseplate. TagA binds to a star-shaped protein to regulate the growth of the contractile sheath.⁶⁷ The NCBI-CDD server and Pfam both predicted that the ImpE superfamily domain contains 1 to 273 amino acid residues. Additionally, the NCBI-CDD server assigned an *E*-value of 1.38e-119 to this domain. Using a motif server like Pfam, 2 additional motifs were found:

Table 2. Ten homologous proteins were chosen from the non-redundant BLASTp database due to their high similarity to our WP_004386962.1 query sequence.

ACCESSION NO.	ORGANISM	PROTEIN NAME	SCORE	PERCENT IDENTITY	E-VALUE
WP_004386962.1	<i>Cronobacter sakazakii</i>	Type VI secretion system accessory protein TagJ	559	100	0.0
WP_105653240.1	<i>Cronobacter malonaticus</i>	Type VI secretion system accessory protein TagJ	558	99.64	0.0
WP_063265155.1	<i>Cronobacter</i>	MULTISPECIES: type VI secretion system accessory protein TagJ	558	99.64	0.0
WP_104671789.1	<i>Cronobacter</i>	MULTISPECIES: type VI secretion system accessory protein TagJ	558	99.64	0.0
WP_076728080.1	<i>Cronobacter sakazakii</i>	Type VI secretion system accessory protein TagJ	558	99.64	0.0
WP_161584471.1	<i>Cronobacter malonaticus</i>	Type VI secretion system accessory protein TagJ	557	99.28	0.0
WP_105622741.1	<i>Cronobacter sakazakii</i>	Type VI secretion system accessory protein TagJ	556	99.28	0.0
ELY2789108.1	<i>Cronobacter sakazakii</i>	Protein of avirulence locus ImpE	556	99.64	0.0
EGT4277351.1	<i>Cronobacter sakazakii</i>	Protein of avirulence locus ImpE	555	99.64	0.0
EJC1154322.1	<i>Cronobacter sakazakii</i>	Protein of avirulence locus ImpE	555	99.28	0.0

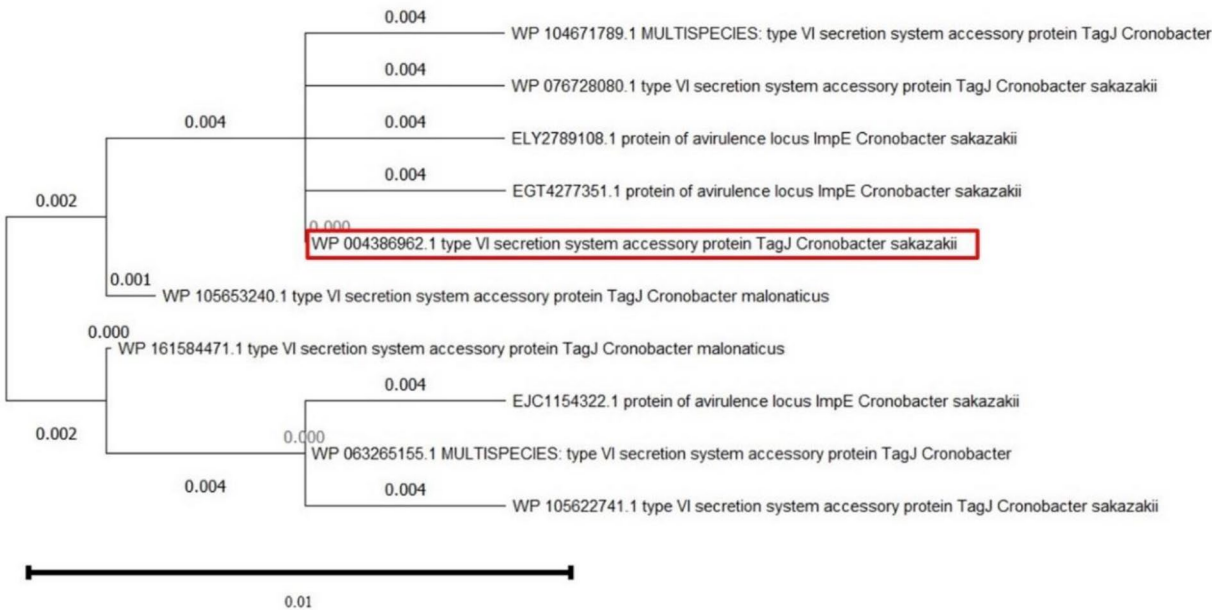


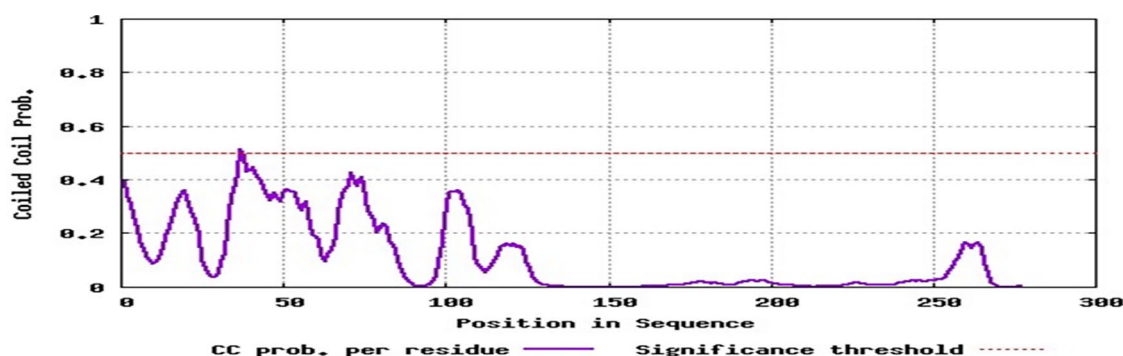
Figure 2. A phylogenetic tree illustrating the ancestral connection of the target protein (boxed) to other TagJ proteins. The phylogenetic tree revealed that 4 proteins, WP_104671789.1, WP_076728080.1, ELY2789108.1, and EGT4277351.1, are closely related to our query (WP_004386962.1) with branch distances of 0.004, respectively.

Tetratricopeptide repeat (PF14559) and ImpE protein (PF07024; Table 3). Tetratricopeptide repeat (TPR) motifs have been identified in a wide range of species from microbes to humans. Numerous biological activities, including nuclear and peroxisomal protein transport, transcriptional regulation,

cell cycle regulation, neurogenesis, and protein folding, are mediated by proteins that contain TPRs.⁶⁸ A “DNA-binding 3-helical bundle” fold was discovered in the protein sequence using the PFP-FunDSeqE method for protein fold pattern recognition. The helical contact of the

Table 3. List of the domains and motifs of the hypothetical protein (accession no: WP_004386962.1).

TOOLS NAME	DOMAIN AND MOTIFS NAME	POSITION (INDEPENDENT E-VALUE)	DESCRIPTION
NCBI-CD search (Domain)	ImpE	1-273 (1.38e-119)	COG4455, ImpE protein
InterProScan (Domain)	TagJ	1-273	IPR009211, Type VI secretion system accessory component TagJ
Pfam (Motifs)	ImpE	147.0.265 (7.6e-40)	PF07024, ImpE protein
Pfam (Motifs)	TPR_19	15.0.73 (0.06)	PF14559, Tetratricopeptide repeat

**Figure 3.** WP_004386962.1 is a putative protein that has undergone functional analysis (coiled-coil interaction). The Y-axis shows the probability of a coiled-coil, while the X-axis shows the position in the sequence.

protein is frequently less regular than a 3-stranded coiled. The formations known as 3-helix bundles are usually single-stranded and have loops connecting the helices. In the tertiary structures of many big proteins, this structural motif is frequently observed as a subdomain. Moreover, lysins, enzyme inhibitors, DNA-binding proteins, and enzymes are among the functionally diverse protein families represented by the 3-helix bundle.⁶⁹ The placements of amino acid numbers in the protein are shown on the X-axis of the coiled-coil graph in Figure 3, while the probability score of the coiled-coil is represented on the Y-axis.

Protein structure prediction

The secondary structure of the protein is predominantly composed of turns, coils, helices, and sheets. According to the SOPMA server analysis, the protein contains a significant proportion of alpha helices (48.01%), followed by random coils (33.57%), beta-turns (5.78%), and extended strands (12.64%; Figure 4). A comparable outcome from the PSIPRED server (Supplemental Figure S2) confirmed the accuracy of the previous result. The 3-dimensional structure of this protein was predicted using the SWISS-MODEL server, with the template protein (4uqx.1.A) serving as a guide. The predicted structure shares 38.28% sequence similarity with the template.⁷⁰ The objective of the template protein is 2 Type VI Secretion Classes Differentiated by the Coevolution of the Accessory HsiE Protein, the TssB-TssC Sheath, and the ATPase ClpV. The

bacteria known as *Pseudomonas aeruginosa* is the template protein.⁵⁵ The 3D structure that was produced using SWISS-MODEL is shown in Figure 4.

3D structural analysis and quality evaluation

The YASARA Energy Minimization Server improved the stability of the protein model by lowering its energy from -121848.9 to -158647.1 kJ/mol. Additionally, the final score increased from -1.33 to -0.50 kJ/mol, indicating a more robust protein structure. The quality of the 3D model was evaluated using the PROCHECK, Verify 3D, QMEAN, and ERRAT programs. PROCHECK analysis revealed that 94.0% of the amino acid residues were positioned in the most preferred region of the Ramachandran plot (see Table 4 and Figure 5A). The Verify 3D server confirmed that the model structure functioned as anticipated, with 83.33% of the residues having an average 3D to 1D score ≥ 0.1 . Using the QMEAN tool, the model was placed into the dark gray zone and had an outstanding QMEAN4 value of -0.53 (Figure 5B). ERRAT further forecasted that the protein structure has a good quality, with a quality factor of 95.7198. In Figure 5C, the model is superimposed on top of the template protein (PDB ID: 4uqx.1.A). The 3D model that was suggested by the RMSD value of 0.458 Å was discovered after superimposition in UCSF Chimera. The Z score of the model, which reflects its overall accuracy, can determine how the input structure compares to the typical scores reported for native proteins of a similar shape. Figure 5D

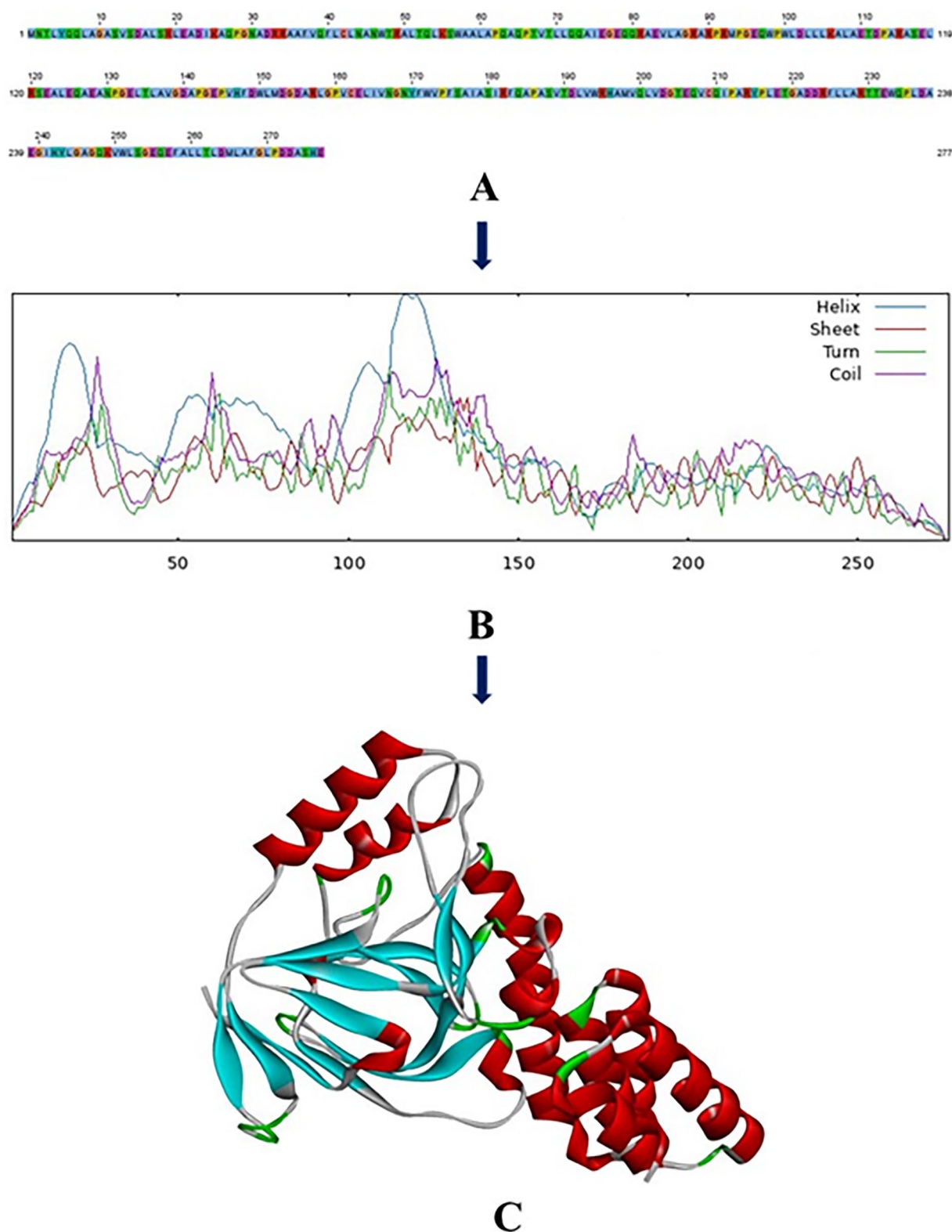


Figure 4. The putative protein WP_004386962.1 shows features reminiscent of ImpE through predictions of its (A) primary, (B) secondary, and (C) 3-dimensional structures. When its sequence of 277 amino acids was first obtained, it was examined to ascertain the makeup of its secondary structure. Significant amounts of alpha helices (48.01%) and extended beta strands (45.05%) were found in the secondary structure anticipating, but beta-turns (5.78%) and random coils (33.57%) were found in smaller proportions: (A) primary structure (NCBI), (B) 2-dimensional structure (SOPMA), and (C) 3-dimensional structure (SWISS-MODEL).

Table 4. Ramachandran plot analysis of the hypothetical protein (WP_004386962.1) provides insights into the conformational preferences of its amino acid residues.

STATISTICS	NO. OF AA RESIDUES	PERCENTAGE (%)
Residues in the most favored regions [A, B, L]	76	93.8
Residues in additional allowed regions [a, b, l, p]	5	6.2
Residues in generously allowed regions [~a, ~b, ~l, ~p]	0	00
Residues in disallowed regions	0	00
No. of non-glycine and non-proline residues	81	100
No. of end residues (excl. Gly and Pro)	2	
No. of glycine residues (shown in triangles)	5	
No. of proline residues	2	
Total no. of residues	90	

and E illustrate that the model generated by ProSA exhibits homology with the template, having Z scores of -7.42 and -7.73 , respectively.

Active site determination

The CASTp web server has shown to be a useful tool for several research projects, such as the evaluation of signaling receptors, the discovery of cancer therapies, the comprehension of drug action mechanisms, and the investigation of issues related to immunological disorders.⁵⁴ This CASTp server was used to analyze the active site and amino acid residues of a model structure (Figure 5F). Identifying and characterizing active site residues is crucial for drug or inhibitor design. Medication or antagonist design depends on the assessment and characterization of active site residues. CASTp forecast states that the model protein's active residues (of 1 biggest active pocket with solvent accessible [SA] area of 120.790 and Volume [SA] area 64.195, respectively) were found to be Trp¹⁰⁰, Trp¹⁰⁰, Leu¹²⁴, Leu¹²⁴, Leu¹²⁴, Leu¹²⁴, Glu¹²⁵, Ala¹²⁷, Ala¹²⁷, Glu¹²⁸, Glu¹²⁸, Ala¹²⁹, Asn¹³⁰, Asn¹³⁰, Asn¹³⁰, Asn¹³⁰, Phe¹⁴⁸, Trp¹⁵⁰, Trp¹⁵⁰, Trp¹⁵⁰, Trp¹⁵⁰, Trp¹⁵⁰, Leu¹⁵¹, Leu¹⁵¹, Met¹⁵², Met¹⁵², Pro¹⁶⁰, Pro¹⁶⁰, Pro¹⁶⁰, Pro¹⁶⁰, Phe¹⁷⁵, Phe¹⁷⁵, Ser¹⁷⁶, Ser¹⁷⁶ consist of the target protein's ImpE superfamily domain in line with the predictions of Pfam, InterProScan, and NCBI-CD Search (described in the "Protein family and phylogeny analysis" section).

Protein-protein interactions analysis

Many predicted interaction partners of the desired protein were found using the STRING database (Figure 6). TssB-2, tssC, AKE93833.1, tssE, AKE93842.1, AKE93825.1, tssB, and AKE93824.1 were the proteins that interacted with each other. These cellular locations and molecular functions of a protein

were found through literature mining and are listed in Table 5. Located in the cytoplasm, the majority of the proteins interact with our hypothetical target protein. The proteins are mostly associated with the supramolecular bacterial complex known as the type VI secretion system (T6SS), which looks like phage tails. When the TssBC sheath transcriptional regulatory system contracts, the toxin delivery system discharges poisons into target cells.⁶⁷ This is crucial to our understanding of the putative protein's identity as the avirulence locus ImpE protein. The lowest (0.15), medium (0.40), high (0.70), and highest (0.90) default thresholds are the ones that STRING advises choosing. Only the edges with a score equal to or higher than the designated threshold are included in the construction of an unweighted network for each of these criteria, creating a network of interconnected proteins.⁷¹ Every protein that interacts has a score of more than 0.70, indicating a strong interaction.

Molecular docking and binding affinity analysis

A docking investigation between the ligands and the model was carried out using the Auto Dock Vina. Hence, there were 6 ligands docked with both the *C. sakazakii* model protein (WP_004386962.1) and the template *Pseudomonas aeruginosa* PAO1 (Q9I746.1) protein. Each ligand had a high binding affinity with both proteins. The binding affinities of the ligands for the model and template proteins ranged from -6.6 to -8.0 kcal/mol. Several interacting residues in the active site were discovered to be the same in both proteins. The results are also in line with the active site assumption of CASTp. Figures 7 and 8, and Table 6 illustrate the binding interactions of 6 compounds within a specific protein pocket. CID: 43507 engages in hydrogen bonding with Ala¹⁰⁷, Ala¹⁰⁹, Glu¹¹⁰, Asp¹¹², Asp²³⁷, and Ala²³⁸, while forming hydrophobic bonds with Pro¹¹³, Ala¹¹⁶, Pro²³⁵, and His²⁴². CID: 5479530 is capable of forming hydrogen bonds with Arg⁸⁸, Leu¹⁰⁸, Thr¹¹¹, Pro²¹⁶, Thr²¹⁹, and Asp²³⁷,

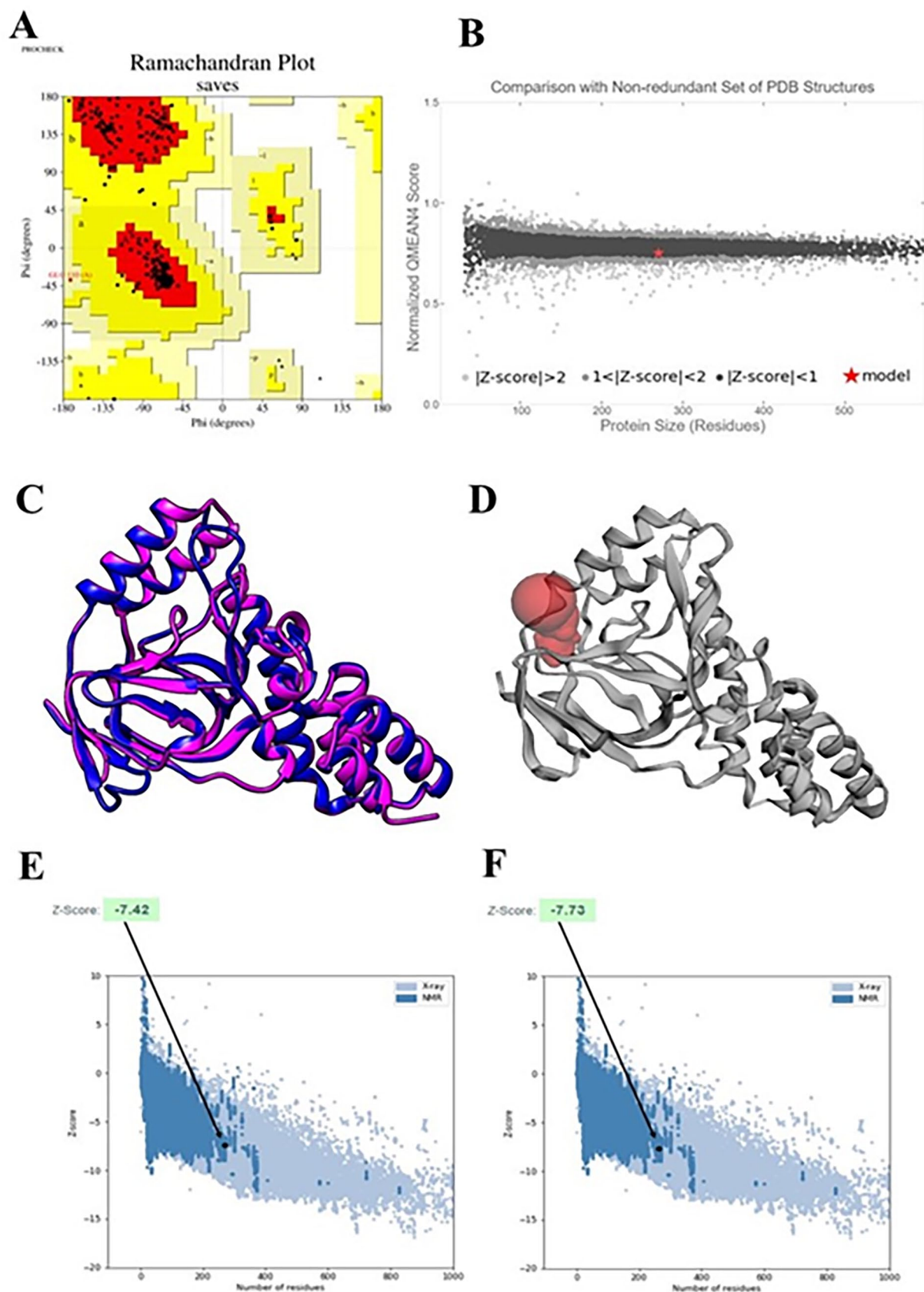


Figure 5. The quality of the model was assessed using 2 methods: (A) a Ramachandran plot generated by the PROCHECK tool to evaluate the structural accuracy; and (B) a QMEAN plot to compare the model's structure with experimental structures of similar size. The QMEAN plot indicated a strong correlation between the model and experimental structures. (C) Using the UCSF Chimera software, (D) A CASTp server-based graphical depiction of an active site, superimpose the template and the model protein (model: royal blue, template: magenta). Z scores of the ProSA server for the target (E) and template (F) modeled protein. The 2 structures were within the usual range for similarly sized native proteins that have been experimentally characterized (NMR and X-ray).

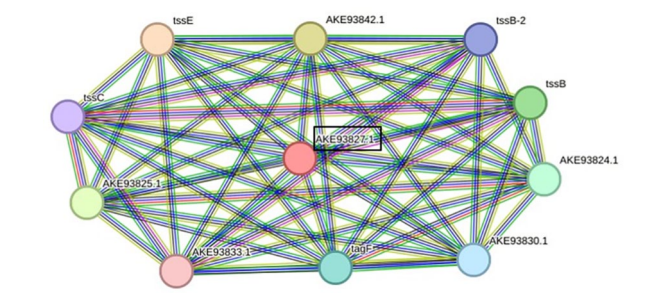


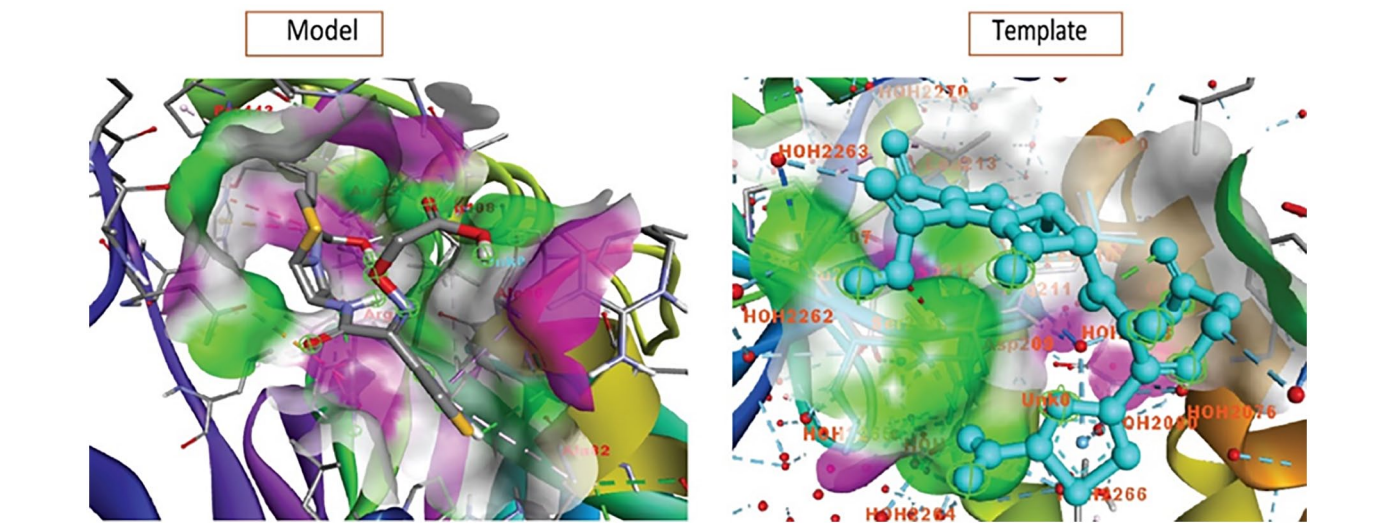
Figure 6. Network of protein-protein interactions for the hypothetical protein from the STRING server. The query proteins are depicted by colored nodes, while the second shell of interactors is represented by white nodes. Proteins with unknown 3D structures are indicated by empty nodes, and those with known or predicted 3D structures are represented by filled nodes.

and engages in hydrophobic interactions with Ala⁸², Ala⁸⁶, Ala¹¹⁶, Arg¹²⁰, Leu²¹⁷, and His²⁴². CID: 47499 primarily binds via hydrogen bonds to Ala⁸⁶, Leu¹⁰⁸, and Pro²³⁵, while also forming hydrophobic interactions with Arg¹²⁰, Asp²³⁷, and Ala¹¹⁶. CID: 5362065 forms hydrogen bonds with Ala⁸², Leu¹⁰⁸, Arg¹²⁰, and Arg¹⁵⁷, and hydrophobic interactions with Ala⁸², Ala⁸⁶, and Ala¹¹⁶. CID: 51081 generates hydrogen bonds with Leu⁸⁵, Leu¹⁰⁸, Glu¹¹⁰, and Thr¹¹¹, and a hydrophobic bond with Ala⁸⁶. Lastly, CID: 37768 establishes hydrogen bonds with Ala⁸², Ala⁸⁶, Leu¹⁰⁸, Thr¹¹¹, Arg¹²⁰, Asp²³⁷, Ala²³⁸, and Glu²³⁹, while interacting hydrophobically with Ala¹⁰⁷.

After the structural and functional annotation of the hypothetical protein was completed successfully, the comparative genomics approach was used to better define our target protein. A BLASTp search was conducted to identify human proteins

Table 5. Analysis of the interacting partners of the target protein (AKE93827.1).

INTERACTING PROTEINS	FUNCTION	LOCATION	SCORE
AKE93827.1	Protein of avirulence locus ImpE	Cytoplasmic	
tssE	VI_zyme domain-containing protein	Cytoplasmic	0.988
AKE93842.1	Type VI secretion system OmpA/MotB family protein	Cytoplasmic	0.976
AKE93825.1	Type VI secretion protein ImpG	Cytoplasmic	0.976
tssB	VI_chp_5 domain-containing protein	Cytoplasmic	0.967
AKE93824.1	Type VI secretion protein	Cytoplasmic	0.956
tagF	Type VI secretion-associated protein	Cytoplasmic	0.940
AKE93830.1	VI_FHA domain-containing protein.	Cytoplasmic	0.934
tssB-2	Putative type VI secretion protein.	Cytoplasmic	0.908
tssC	EvpB family type VI secretion protein.	Cytoplasmic	0.894



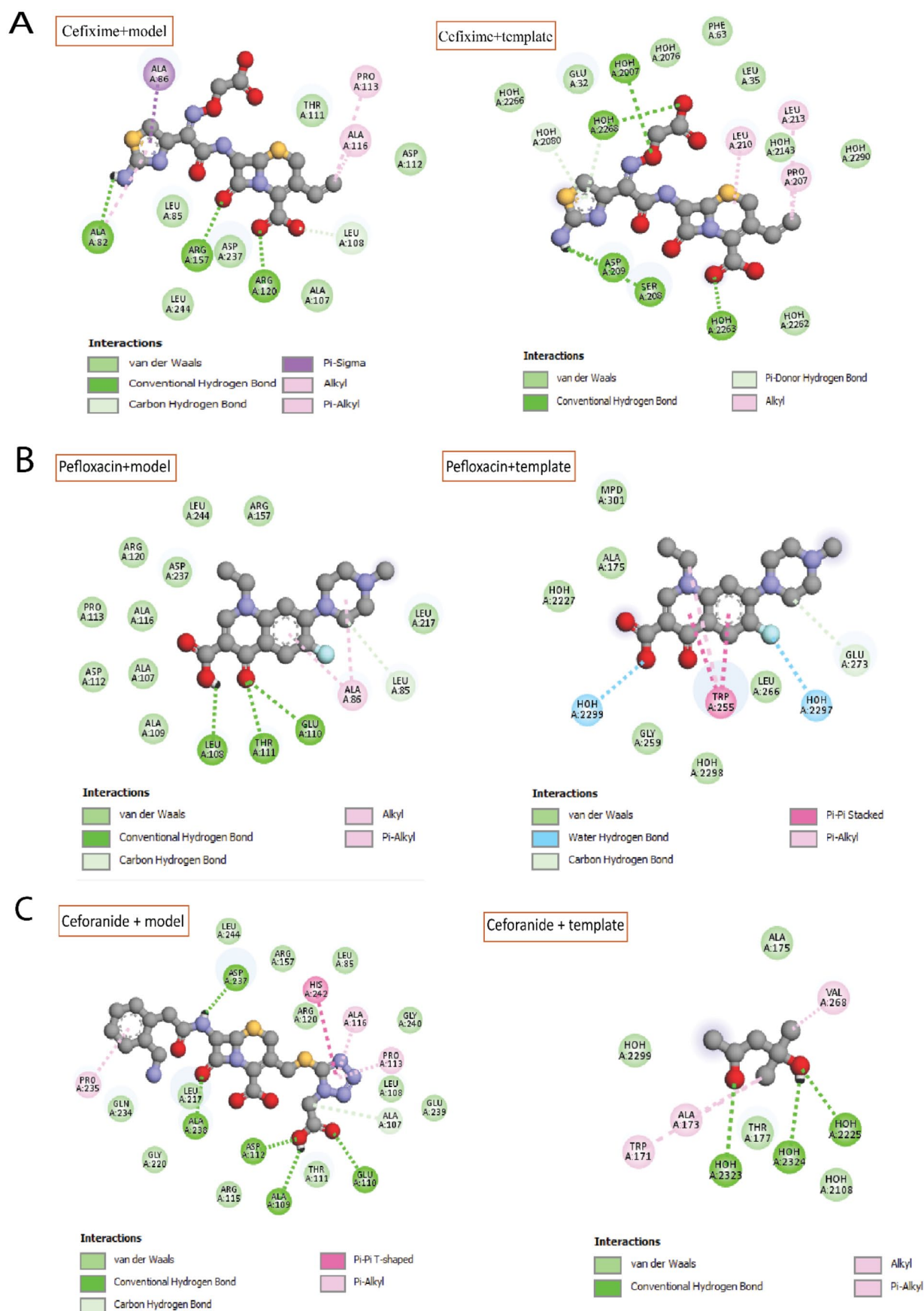


Figure 8. (Continued)

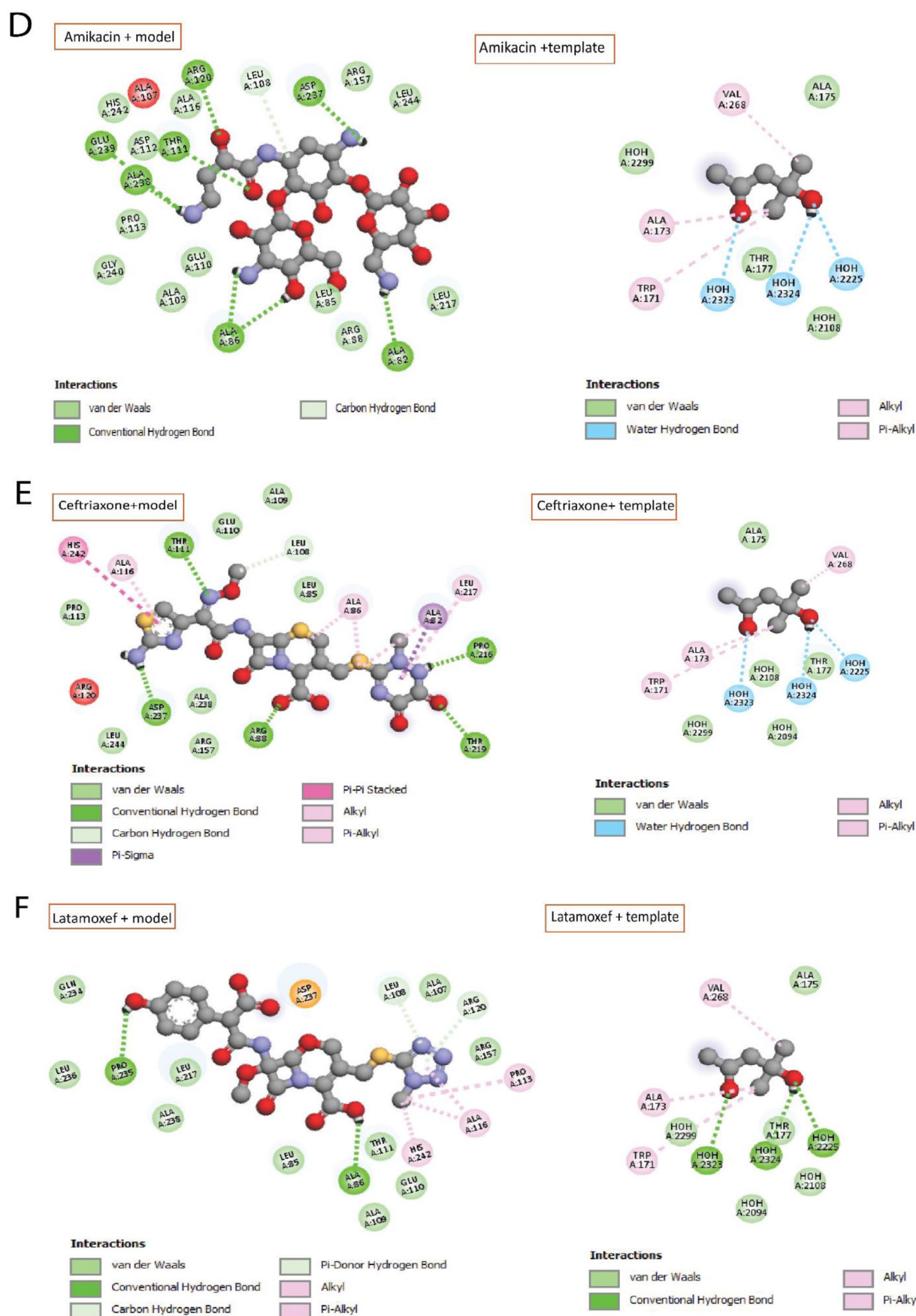


Figure 8. Key interacting residues of hypothetical protein (WP_004386962.1) model and also a template (Q9I746.1) with: (A) Cefixime, (B) Pefloxacin, (C) Ceforanide, (D) Amikacin, (E) Ceftriaxone, and (F) Latamoxef antibiotics.

Table 6. Interacting residues of target protein and model using Auto Dock Vina.

TARGET PROTEIN	LIGANDS (CID)	BINDING AFFINITY (KCAL/MOL)	INTERACTING RESIDUES	
			HYDROGEN BOND	HYDROPHOBIC BOND
Model (WP_004386962.1)	Ceforanide (43507)	−7.5	Ala ¹⁰⁷ , Ala ¹⁰⁹ , Glu ¹¹⁰ , Asp ¹¹² , Asp ²³⁷ , Ala ²³⁸	Pro ¹¹³ , Ala ¹¹⁶ , Pro ²³⁵ , His ²⁴²
	Ceftriaxone (5479530)	−7.4	Arg ⁸⁸ , Leu ¹⁰⁸ , Thr ¹¹¹ , Pro ²¹⁶ , Thr ²¹⁹ , Asp ²³⁷	Ala ⁸² , Ala ⁸⁶ , Ala ¹¹⁶ , Arg ¹²⁰ , Leu ²¹⁷ , His ²⁴²
	Latamoxef (47499)	−7.1	Ala ⁸⁶ , Leu ¹⁰⁸ , Pro ²³⁵	Arg ¹²⁰ , Asp ²³⁷ , Ala ¹¹⁶
	Cefixime (5362065)	−6.9	Ala ⁸² , Leu ¹⁰⁸ , Arg ¹²⁰ , Arg ¹⁵⁷	Ala ⁸² , Ala ⁸⁶ , Ala ¹¹⁶
	Pefloxacin (51081)	−6.7	Leu ⁸⁵ , Leu ¹⁰⁸ , Glu ¹¹⁰ , Thr ¹¹¹	Ala ⁸⁶
	Amikacin (37768)	−6.6	Ala ⁸² , Ala ⁸⁶ , Leu ¹⁰⁸ , Thr ¹¹¹ , Arg ¹²⁰ , Asp ²³⁷ , Ala ²³⁸ , Glu ²³⁹	Ala ¹⁰⁷
Template (Q91746.1)	Ceforanide (43507)	−7.2	Phe ⁶³ , Ser ²⁰⁸	Glu ³²
	Ceftriaxone (5479530)	−7.8	Ser ²⁰⁸	Leu ²¹³
	Latamoxef (47499)	−7.9		Glu ³² , Phe ⁶³
	Cefixime (5362065)	−7.5	Ser ²⁰⁸ , Asp ²⁰⁹	Pro ²⁰⁷ , Leu ²¹⁰ , Leu ²¹³
	Pefloxacin (51081)	−8.0	Glu ²⁷³	Trp ²⁵⁵
	Amikacin (37768)	−7.5	Ser ²⁰⁸ , Thr ⁵⁹	

similar to the target protein. The target protein was identified as a novel *C. sakazakii* protein after no resemblance was found to any known human protein. To reduce side effects, a good therapy option would target microbial proteins that are not comparable to human proteins.

Discussion

The increasing affordability of sequencing technology has led to a surge in genomics and proteomics data. However, research on hypothetical proteins has lagged. Studying hypothetical proteins could provide valuable insights into bacterial metabolism, disease progression, drug discovery, and disease prevention strategies.⁷² Using several bioinformatics resources, this study structurally and functionally described the putative protein WP_004386962.1 from the *Cronobacter sakazakii* 7G strain. The protein analyzed had 277 amino acids, a molecular weight of 30427.40 Da, a theoretical pI of 4.54, and a GRAVY of −0.140 (Table 1). These characteristics suggest a soluble protein. The CELLO server predicted that this protein is localized within the cytoplasm. The secondary structure of the protein consists of an alpha helix, a beta-turn, a random coil, and an extended strand, with the random coil being a significant prevalent. Based on domain and motif analysis, we predicted that the desired putative protein would be an ImpE superfamily protein, which is categorized with high confidence by all annotation methods as TagJ (also known as HsiE1) of T6SS (Table 3). The BLASTp analysis, which compared the sequence against a nonredundant database, revealed a high degree of sequence similarity (up to 99%) with various T6SS TagJ proteins. This strong similarity further

confirms the earlier prediction. TagJ (HsiE1) is an auxiliary protein in the type VI secretion system (T6SS) of bacteria. It contributes to TssC/TssB and ClpV the sheath's interaction.⁵⁵ The type VI secretion system (T6SS), like phage tails, is a bacterial supramolecular group. It is a toxin delivery system that injects poisons into target cells when the TssBC sheath contracts.⁶⁶ The Type VI secretion system (T6SS) increases the pathogenicity of dangerous bacteria like *Pseudomonas aeruginosa*, *Escherichia coli*, and *Vibrio cholerae*. This increases the formation of biofilm, protects the host immune system, and increases antibiotic resistance. Many strains of *E. coli* are known for their significant drug resistance. T6SS consists of 2 tubular structures, TssB and TssC, which are disassembled by ClpV. TagJ/HsiE is commonly associated with one of these components.^{55,73} The 3D structure of a protein, obtained via the SWISS-MODEL service, completed various model quality assurance tests, including PROCHECK, QMEAN, Verify 3D, and ERRAT. The YASARA energy minimization method enhanced the reliability of the 3D structure. When the hypothetical protein was compared to the *Pseudomonas aeruginosa* potential cytoplasmic protein (PDB ID: 4UQX) using UCSF Chimera, the 3D structures aligned well with a root-mean-square deviation (RMSD) of 0.458 Å (explained in this “Structure analysis and model quality assessment” section). The RMSD value indicates high sequence identity, suggesting minimal variability in protein structures. Our result, approaching zero, indicates a strong similarity between the template structure and the hypothetical model.⁷⁴ The CASTp server calculated active site amino acid residues in the ImpE superfamily domain area, which aligned with functional annotation

tools' predictions. Using the Autodock Vina, molecular docking was used to determine how the 6 ligands interacted with target proteins. According to the docking result, hydrogen and hydrophobic bonds were identified with interacting residues. These interactions are essential for protein folding, structure maintenance, and molecular recognition, thereby stabilizing ligands and enhancing therapeutic efficacy.^{75,76} Our findings were further verified by the observation of a high binding affinity among these ligands and both the target protein and the *Pseudomonas aeruginosa* TagJ protein (Table 6). The active sites of the proteins were shown to have a large number of identical interaction residues. As a second-generation cephalosporin, ceforanide has antibacterial activity against many Enterobacteriaceae.⁷⁷ Ceftriaxone is a popular third-generation cephalosporin antibiotic renowned for its broad-spectrum activity. It is particularly valued for its effectiveness against multidrug-resistant Enterobacteriaceae.⁵⁸ Latamoxef is highly effective against Gram-negative bacteria, including Enterobacteriaceae and *Bacteroides fragilis*. Its potent activity makes it a promising option for treating intra-abdominal infections in immunocompromised patients, as well as neonatal Gram-negative bacillary meningitis, due to its strong efficacy against Gram-negative bacilli and user-friendly administration.⁵⁷ Cefixime is a third-generation, semisynthetic cephalosporin, and an oral broad-spectrum antibiotic.⁷⁸ Cephalosporins, which belong to the beta-lactam antibiotic class, can be used to treat meningitis, resistant bacterial strains, skin infections, and inflammation induced by Gram-positive and Gram-negative bacteria. Pefloxacin effectively targets a wide range of bacteria, including both Gram-positive and Gram-negative types, making it a versatile antibiotic.⁷⁹ Amikacin is particularly efficient against more resistant Gram-negative bacteria and inhibits a lot of aerobic Gram-negative bacteria in the Enterobacteriaceae family. The 6 ligands used in this study have antibacterial activity against a wide range of Gram-negative bacteria, from those of the Enterobacteriaceae family. Given that our chosen hypothetical protein from *C. sakazakii* is also a Gram-negative bacterium within the Enterobacteriaceae family, these inhibitors represent promising options for treating this bacterial infection. We have clarified that, due to the lack of experimental validation, the *in silico* characterization of the hypothetical protein may not fully reflect its biological complexity. TagJ (HsiE1), an auxiliary component of the type VI secretion system (T6SS), is a known virulence factor, but its role in *Cronobacter* species is unknown.⁸⁰ The emphasis on a single putative protein may overshadow other critical proteins and the significance of the TagJ-associated type VI secretion system (T6SS), limiting our understanding of the metabolic flexibility of *C. sakazakii*. The limited nature of this study emphasizes the crucial importance of further investigating the metabolic pathways of the organism. Developing possible therapeutics for this bacterium infection will require additional molecular study utilizing both *in-vivo* and *in-vitro* models.

Conclusions

This study focuses on identifying a hypothetical protein (WP_004386962.1) associated with the type VI secretion system (T6SS) in *C. sakazakii* utilizing *in-silico* methods and molecular docking. The research highlights the protein's sub-cellular localization, functional domains, motifs, and docking interactions with various ligands, all of which are crucial for understanding the pathogenesis of *C. sakazakii*. The characteristics of a protein, including its role in drug resistance, are investigated to improve our understanding of *C. sakazakii* pathophysiology and to contribute in medication discovery. Six FDA-approved inhibitor compounds were tested against this hypothetical protein, with Ceforanide showing (−7.5 kcal/mol) the most promising binding interaction score. Further research is required to validate these findings. Furthermore, a comparative genomics investigation identified a unique protein in *C. sakazakii* as a possible therapeutic target. The T6SS is instrumental in biofilm formation, which increases bacterial antibiotic resistance and complicates infection treatment. Further studies and experimental validations are necessary to fully comprehend T6SS and its effectors, supporting the development of effective treatment strategies for this pathogenic bacterium.

Acknowledgements

We want to express our gratefulness to the members of the Center for Advanced Bioinformatics and Artificial Intelligence Research Lab, led by Dr. Md Habibur Rahman, for their assistance and insightful contributions to the study.

Author Contribution Statement

Nurun Nahar Akter: Provided concept, experimenting, and built the paper; Md. Moin Uddin: Data analysis, edited and reviewed the manuscript; Israt Jahan Asha, Md Soyeb Uddin, and Fahadul Alam: Analyzed and interpreted the data; Md. Arju Hossain: Edited and reviewed the manuscript; Nesar Uddin, Siratul Kubra Shifat, and Md. Abu Zihad: Data analysis and formal analysis; Md Habibur Rahman: Designed the experiments and supervised the whole project.

Data Availability

The corresponding author can provide the data that was utilized to support the study upon request.

Ethical Statement

No ethical approval was required for this manuscript.

ORCID iDs

Nurun Nahar Akter  <https://orcid.org/0009-0003-4730-3615>

Md. Moin Uddin  <https://orcid.org/0009-0005-4282-2330>

Israt Jahan Asha  <https://orcid.org/0009-0009-1901-3139>

Md Habibur Rahman  <https://orcid.org/0000-0002-5068-2690>

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

- Yan X, Gurtler JB. *Cronobacter* (Enterobacter) sakazakii. In: Batt CA, Tortorello ML, eds. *Encyclopedia of Food Microbiology*. 2nd ed. Elsevier; 2014:528-532.
- Schoch CL, Ciufu S, Domrachev M, et al. NCBI taxonomy: a comprehensive update on curation, resources and tools. *Database*. 2020;2020:baaa062.
- Lai KK. Enterobacter sakazakii infections among neonates, infants, children, and adults. Case reports and a review of the literature. *Medicine*. 2001;80:113-122.
- Srikumar S, Cao Y, Yan Q, et al. RNA sequencing-based transcriptional overview of xerotolerance in *Cronobacter sakazakii* SP291. *Appl Environ Microbiol*. 2019;85:e01993. doi:10.1128/AEM.01993-18
- Hunter CJ, Petrosyan M, Ford HR, Prasadarao NV. Enterobacter sakazakii: an emerging pathogen in infants and neonates. A review of their virulence and environmental-adaptive traits. *J Med Microbiol*. 2014;63:1023-1037.
- Kalyantanda G, Shumyak L, Archibald LK. *Cronobacter* species contamination of powdered infant formula and the implications for neonatal health. *Front Pediatr*. 2015;3:56. doi:10.3389/fped.2015.00056
- Kim K, Kim KP, Choi J, et al. Outer membrane proteins A (OmpA) and X (OmpX) are essential for basolateral invasion of *Cronobacter sakazakii*. *Appl Environ Microbiol*. 2010;76:5188-5198.
- Wyllie R, Hyams JS, Kay M. Pediatric gastrointestinal and liver disease E-book. *Elsevier Health Sci*. 2020;24:978-1001.
- Mullane NR, Iversen C, Healy B, et al. Enterobacter sakazakii an emerging bacterial pathogen with implications for infant health. *Minerva Pediatr*. 2007;59:137-148.
- Skovgaard N. New trends in emerging pathogens. *Int J Food Microbiol*. 2007;120:217-224.
- Gurtler JB, Kornacki JL, Beuchat LR. Enterobacter sakazakii: a coliform of increased concern to infant health. *Int J Food Microbiol*. 2005;104:1-34.
- Uddin R, Saeed K. Identification and characterization of potential drug targets by subtractive genome analyses of methicillin resistant *Staphylococcus aureus*. *Comput Biol Chem*. 2014;48:55-63.
- Iversen C, Forsythe S. Risk profile of Enterobacter sakazakii, an emergent pathogen associated with infant milk formula. *Trends Food Sci Technol*. 2003;14:443-454.
- Carvalho GG, Calarga AP, Teodoro JR, et al. Isolation, comparison of identification methods and antibiotic resistance of *Cronobacter* spp. in infant foods. *Food Res Intern*. 2020;137:109643.
- Depardieu F, Podglajen I, Leclercq R, Collatz E, Courvalin P. Modes and modulations of antibiotic resistance gene expression. *Clin Microbiol Rev*. 2007;20:79-114.
- Lin X, Li X, Lin X. A review on applications of computational methods in drug screening and design. *Molecules*. 2020;25:1375. doi:10.3390/molecules25061375
- Barh D, Tiwari S, Jain N, et al. In silico subtractive genomics for target identification in human bacterial pathogens. *Drug Dev Res*. 2011;72:162-177.
- Hasan M, Azim KF, Begum A, et al. Vaccinomics strategy for developing a unique multi-epitope monovalent vaccine against Marburg marburgvirus. *Infect Genet Evol*. 2019;70:140-157.
- Choi HP, Juarez S, Ciordia S, et al. Biochemical characterization of hypothetical proteins from *Helicobacter pylori*. *PLoS One*. 2013;8:e66605. doi:10.1371/journal.pone.0066605
- Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics*. 2008;92:255-264.
- Hoskeri H J. Functional annotation of conserved hypothetical proteins in rickettsia massiliae MTU5. *J Comput Sci Syst Biol*. 2010;03:050-052. doi:10.4172/jcsb.1000055
- Shahbaaz M, Hassan MI, Ahmad F. Functional annotation of conserved hypothetical proteins from *Haemophilus influenzae* rd KW20. *PLoS One*. 2013;8:e84263. doi:10.1371/journal.pone.0084263
- Shahbaaz M, Bisetty K, Ahmad F, Hassan MI. Current advances in the identification and characterization of putative drug and vaccine targets in the bacterial genomes. *Curr Top Med Chem*. 2016;16:1040-1069.
- Benson DA, Cavanaugh M, Clark K, et al. GenBank. *Nucleic Acids Res*. 2013;41:D41-D42.
- Wilkins MR, Gasteiger E, Bairoch A, et al. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*. 1999;112:531-552.
- Garg VK, Avasthi H, Tiwari A, et al. MFPP1 - multi FASTA ProtParam interface. *Bioinformatics*. 2016;12:74-77. doi:10.6026/97320630012074
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403-410.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35:1547-1549.
- Edgar RC, Batzoglou S. Multiple sequence alignment. *Curr Opin Struct Biol*. 2006;16:368-373.
- Yu C, Chen Y, Lu C, Hwang J. Prediction of protein subcellular localization. *Proteins*. 2006;64:643-651.
- Yu NY, Wagner JR, Laird MR, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*. 2010;26:1608-1615.
- Bhasin M, Garg A, Raghava GP. PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics*. 2005;21:2522-2524.
- Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*. 1998;14:378-379.
- Lu S, Wang J, Chitsaz F, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. 2020;48:D265-D268.
- Quevillon E, Silventoinen V, Pillai S, et al. InterProScan: protein domains identifier. *Nucleic Acids Res*. 2005;33:W116-W120.
- Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res*. 2002;30:281-283.
- El-Gebali S, Mistry J, Bateman A, et al. The pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47:D427-D432.
- Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res*. 2006;34:W369-W373.
- Ludwiczak J, Winski A, Szczepaniak K, Alva V, Dunin-Horkawicz S. DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics*. 2019;35:2790-2795.
- Shen HB, Chou KC. Predicting protein fold pattern with functional domain and sequential evolution information. *J Theor Biol*. 2009;256:441-446.
- Buchan DWA, Jones DT. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res*. 2019;47:W402-W407.
- Combet C, Blanchet C, Geourjon C, Deléage G. NPS@: network protein sequence analysis. *Trends Biochem Sci*. 2000;25:147-150.
- Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46:W296-W303.
- Krieger E, Joo K, Lee J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins*. 2009;77 Suppl 9:114-122.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 1993;26:283-291.
- Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol*. 1997;277:396-404.
- Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*. 2011;27:343-350.
- Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. 1993;2:1511-1519.
- Pettersen EF, Goddard TD, Huang CC, et al. UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605-1612.
- 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:W407-W410.
- Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47:D607-D613.
- Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res*. 2018;46:W363-W367.
- Förster A, Planamente S, Manoli E, Lossi NS, Freemont PS, Filloux A. Coevolution of the ATPase ClpV, the sheath proteins TssB and TssC, and the accessory protein TagJ/HsiE1 distinguishes type VI secretion classes. *J Biol Chem*. 2014;289:33032-33043.
- Rose PW, Beran B, Bi C, et al. The RCSB protein data bank: redesigned web site and web services. *Nucleic Acids Res*. 2011;39:D392-D401.
- Carmine AA, Brogden RN, Heel RC, Romankiewicz JA, Speight TM, Avery GS. Moxalactam (latamoxef). A review of its antibacterial activity, pharmacokinetic properties and therapeutic use. *Drugs*. 1983;26:279-333.

58. Richards DM, Heel RC, Brogden RN, Speight TM, Avery GS. Ceftriaxone. A review of its antibacterial activity, pharmacological properties and therapeutic use. *Drugs*. 1984;27:469-527.
59. Kim S, Thiessen PA, Bolton EE, et al. PubChem substance and compound databases. *Nucleic Acids Res*. 2016;44:D1202-D1213.
60. Rosignoli S, Paiardini A. Boosting the full potential of pymol with Structural Biology plugins. *Biomolecules*. 2022;12:1764. doi:10.3390/biom12121764
61. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31:455-461.
62. Jejurikar BL, Rohane SH. Drug designing in discovery studio. *Asian J Res Chem*. 2021;14:135-138.
63. Wei L, Liu Y, Dubchak I, Shon J, Park J. Comparative genomics approaches to study organism similarities and differences. *J Biomed Inform*. 2002;35:142-150.
64. Choudhuri S (ed.). Chapter 6—sequence alignment and similarity searching in genomic databases: BLAST and FASTA. In *Bioinformatics for Beginners, Sequence Alignment and Similarity Searching in Genomic Databases*. Elsevier Academic Press; 2014:133-155.
65. Chen L, Zou Y, She P, Wu Y. Composition, function, and regulation of T6SS in *Pseudomonas aeruginosa*. *Microbiol Res*. 2015;172:19-25.
66. Planamente S, Salih O, Manoli E, Albesa-Jové D, Freemont PS, Filloux A. TssA forms a gp6-like ring attached to the type VI secretion sheath. *EMBO J*. 2016;35:1613-1627.
67. Liebl D, Robert-Genthon M, Job V, Cogoni V, Attrée I. Baseplate component TssK and spatio-temporal assembly of T6SS in *Pseudomonas aeruginosa*. *Front Microbiol*. 2019;10:1615. doi:10.3389/fmicb.2019.01615
68. D'Andrea LD, Regan L. TPR proteins: the versatile helix. *Trends Biochem Sci*. 2003;28:655-662.
69. Efimov AV. Favoured structural motifs in globular proteins. *Structure*. 1994; 2:999-1002.
70. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res*. 2003;31:3381-3385.
71. Bozhilova LV, Whitmore AV, Wray J, Reinert G, Deane CM. Measuring rank robustness in scored protein interaction networks. *BMC Bioinformatics*. 2019;20:446. doi:10.1186/s12859-019-3036-6
72. Sen T, Verma NK. Functional annotation and curation of hypothetical proteins present in A newly emerged serotype 1c of *Shigella flexneri*: emphasis on selecting targets for virulence and vaccine design studies. *Genes*. 2020; 11:340.
73. Cianfanelli FR, Monlezun L, Coulthurst SJ. Aim, Load, fire: the Type VI secretion system, a bacterial nanoweapon. *Trends Microbiol*. 2016;24:51-62.
74. Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci*. 2001;10:1470-1473.
75. Patil R, Das S, Stanley A, Yadav L, Sudhakar A, Varma AK. Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. *PLoS One*. 2010;5:e12029. doi:10.1371/journal.pone.0012029
76. Hubbard RE, Kamran Haider M. Hydrogen bonds in proteins: role and strength. *Encyclopedia of Life Sciences*. 2010;1:1-6.
77. Campoli-Richards DM, Lackner TE, Monk JP. Ceforanide. A review of its antibacterial activity, pharmacokinetic properties and clinical efficacy. *Drugs*. 1987;34:411-437.
78. Brogden RN, Campoli-Richards DM. Cefixime. A review of its antibacterial activity, pharmacokinetic properties and therapeutic potential. *Drugs*. 1989; 38:524-550.
79. Bressolle F, Gonçalves F, Gouby A, Galtier M. Pefloxacin clinical pharmacokinetics. *Clin Pharmacokinet*. 1994;27:418-446.
80. Wang M, Cao H, Wang Q, Xu T, Guo X, Liu B. The roles of two Type VI secretion systems in *Cronobacter sakazakii* ATCC 12868. *Front Microbiol*. 2018; 9:2499. doi:10.3389/fmicb.2018.02499