Research article

Islam, S., et al., In-silico functional annotation of a hypothetical protein from *Edwardsiella tarda* revealed Proline metabolism and apoptosis in fish. International Journal of Life Sciences and Biotechnology, 2022. 5(1): p. 78-96. DOI: 10.38001/ijlsb.1032171

In-silico functional annotation of a hypothetical protein from *Edwardsiella tarda* revealed Proline metabolism and apoptosis in fish

Sk Injamamul Islam[1]*, Saloa Sanjida[2], Moslema Jahan Mou [3], Sarower Mahfuj [1], Saad Nasir[4]

ABSTRACT

Edwardsiella tarda is one of the most widespread pathogens in aquatic species. A wide variety of diseases can be caused by this microbe, including Edwardsiella septicaemia but clinical signs of infection differ between species of fish. The fact that the bacteria is resistant to a wide range of antimicrobials is extremely important. Furthermore, several proteins in its genome are classified as hypothetical proteins (HPs). As a result, the current work sought to elucidate the roles of an HP found in the genome of *E.tarda*. To determine the structure and function of this protein, many bioinformatics methods were used. To locate the homologous protein, the sequence similarity was searched across the available bioinformatics databases. Quality evaluation methods were used to predict and confirm the secondary and tertiary structure. Additionally, the active site and interacting proteins were examined using CASTp and the STRING server. An important biological activity of the HP is that it contains single functional domains that may be responsible for host-cell invasion and autolysis. Further, protein-protein interactions within selected HP revealed several functional partners that are essential for bacterial survival. One such partner is the proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase (putA) of *E. tarda*. In addition, molecular docking and simulation results showed stable bonding between HP and Proline metabolism protein. Finally, the current work shows that the annotated HP is associated with possible mitochondrial metabolism and autolysis formation activities, as well as having a stable binding with the putA protein, which might be of significant relevance to future bacterial genetics research.

ARTICLE HISTORY
Received
5 December 2021
Accepted
23 January 2022

KEYWORDS
Hypothetical protein, molecular docking, putA protein, autolysis, *E. tarda*.

Introduction

Edwardsiella tarda (*E. tarda*) is found in a wide range of hosts, including humans, other animals, and fish. This is worth noting that *E. tarda*-related fish infections cause the most

[1] Department of Fisheries and Marine Bioscience, Faculty of Biological Science, Jashore University of Science and Technology, Jashore-7408, Bangladesh
[2] Department of Environmental Science and Technology, Faculty of Applied Science and Technology, Jashore University of Science and Technology, Jashore-7408, Bangladesh
[3] Department of Genetic Engineering and Biotechnology, Faculty of Earth and Life Science, University of Rajshahi, Bangladesh
[4] Department of clinical medicine and surgery, Faculty of Veterinary Medicine, University of veterinary and animal sciences, Pakistan
* Corresponding Author**:** Sk Injamamul Islam, E-mail: 6378506331@student.chula.ac.th

common bacterial illness in fish and can result in massive infection and mortality in both saline and fresh water. It may be found in a variety of aquatic ecosystems all around the world [1]. Edwardsiellosis and systemic hemorrhagic septicemia, both caused by *E. tarda*, were characterized by internal abscesses and extensive skin inflammation with coloration on fishes [2]. The catastrophic fatalities of farmed fish caused by *E. tarda* infection have resulted in significant financial damage in the United States of America, notably in Japan and India [3]. It seems to be worth noting that the infection in people usually results in dysentery, gastroenteritis, meningitis, delayed wound healing, and sometimes even mortality [4]. In addition to invasion capability, *E. tarda* can release siderophores implicated in induced inflammation and abnormal host cell function, to infect epithelial cells and form biofilms [5-8]. To minimize pathogenic bacterial burdens, tons of antibiotics have been thrown into pathogen-prone areas in recent years. Rather, this approach has worsened the threat posed to the aquaculture industry and public health by drug-resistant strains, such as *E. tarda* [9, 10]. Virulent *E. tarda*, identified from edwardsiellosis-infected Japanese flounder, has shown great resistance to kanamycin, tetracycline, ampicillin, and streptomycin. It also appeared to be resistant to a variety of other antibiotics often used in farmed fish [11]. To a large extent, tetracycline resistance is mediated mainly by proton-dependent efflux pumps, which are found in gram-negative bacteria, particularly *E.tarda*, as well as ribosomal proteins found in gram-positive bacteria [12]. Antibiotic resistance, on the other hand, is a complex phenomenon, with resistance mechanisms presumably involving complex networks of genes, proteins, and biological processes.

Because of the unique fimbriae it creates, *E. tarda* has a strong attraction for red blood cells. *E. tarda* exhibits hemagglutination characteristics as a result of this. The fimbriae are encoded in the genome as a 534-base-pair region [13]. Nevertheless, many of this bacterium's proteins are classified as HPs since their structures and biological activities are unknown. Such proteins can be extremely useful, and their annotation can lead to new insights into their structures, routes, and activities. Consequently, bioinformatics approaches can be utilized to predict and analyze various forms of the structure of those HPs, their biological functions, and their interactions with other proteins.

It got simpler to attribute function to an HP utilizing numerous bioinformatics methods as the in-silico study progressed. We aimed to develop a better understanding of the protein and further drug targets through the assignment of structural and biological functions to hypothetical protein AAW15_09260 (AKH89314.2) of *E.tarda*. Protein-protein interaction was investigated and subcellular distribution, secondary structure, and active site were predicted. In addition, homology modeling techniques were used to attempt to produce a good quality model of the AKH89314.2.

Methods

Identification of similarity and retrieval of sequences

The NCBI database was used to obtain the sequence information for the hypothetical protein (HP) (AKH89314.2). After that, the sequence was saved as a FASTA format and submitted to multiple prediction servers for in-silico analysis. Previously, scientists describe the possible role and effect of this selected HP in channel catfish (*Ictalurus punctatus*), however, they did not functionalize and characterize the HP [14]. Therefore, this HP of *Edwardsiella tarda* needs to be characterized to reveal its role in channel catfish. A similarity search was conducted with the NCBI protein database to provide a first prediction regarding the function of the targeted HP (https://www.ncbi.nlm.nih.gov/) against non-redundant [15] database to use the BLASTp tool to search proteins that may have similar characteristics to the HP [16].

Phylogeny analysis and multiple sequence alignment

Multiple sequence alignments were performed using the BioEdit biological sequence alignment editor between the HP and proteins with similar structural characteristics to the HP [17]. The phylogenetic analysis was performed using an older version of the Molecular Evolutionary Genetic Study (MEGA) (https://megasoftware.net/).

Physiochemical properties analysis

ExPASy's ProtParam (http://web.expasy.org/protparam/) tool was used to determine physical and chemical parameters such as molecular weight, amino acid composition, theoretical pI, instability index, extinction coefficient, atomic composition, estimated half-life, the total number of positively charged residues (Arg + Lys), the total number of negatively charged residues (Asp + Glu), aliphatic index, and grand average of hydropathicity (GRAVY) [18].

Subcellular localization analysis

CELLO anticipated subcellular localization [19]. The results were also compared to PSORTb subcellular localization predictions [20], PSLpred [21], and SOSUIGramN. TMHMM [22], HMMTOP [23], and CCTOP [24] were used for the topology prediction.

Identification of conserved domains, motifs, folds, families, and superfamilies

A search was conducted on the database of conserved domains (CDD, available at NCBI) [25], for the conserved domain. The Motif (Genome Net) server was used to find protein motifs [26]. The evolutionary connections of the protein were assigned using Pfam [27] and SuperFamily [28] database. For the functional analysis of the protein, the protein sequence analysis and classification software InterProScan [29] was used. The PFP-FunD SeqE server [30] was used to recognize protein folding patterns.

Prediction of secondary structure

The secondary structure of proteins was predicted using PSI-blast-based secondary structure prediction (PSIPRED) [31]. Its algorithm employs artificial neural networks and machine learning techniques. It is indeed a server-side application with a front-end website that can predict a protein's secondary structure (beta sheets, alpha helixes, and coils) based on its primary sequence.

3D structure prediction, Refinement, and Validation

The three-dimensional structure of the target protein was predicted using the Raptorx server (http://raptorx.uchicago.edu/) [32]. The protein's 3D structure was refined using GalaxyWeb. In homology modeling, which is based on empirically proven 3D protein structures, the structure's validity is a vital step. The suggested protein model was submitted to ProSA-web for basic confirmation. The z-score, which represents the overall character of the model, was predicted by the server. If the z-scores of the predicted model are outside the scale of the property for local proteins, the structure is incorrect. A Ramachandran plot analysis was performed utilizing the Ramachandran Plot Server to establish the overall quality of the protein (https://zlab.umassmed.edu/bu/rama/).

Assessment of model quality

Subsequently, the predicted three-dimensional structure was evaluated using PROCHECK, Verify3D, and ERRAT Structure Evaluation server.

Protein-Protein Interaction Analysis

Protein functions are determined by interactions between their residues. The STRING database (http://string-db.org/) was employed in this investigation, which analyzes physical and functional correlations to discover known and expected protein interactions. Genomic context, high-throughput investigations, (Conserved) Co-expression and prior knowledge were used to make this decision. This database quantitatively incorporates interaction data from the following sources [33].

Protein disulfide bonds

The formation of disulfide bonds between cysteine residues in a protein is critical for its folding into a functional and stable shape. CYSPRED and DIANNA were used to predict disulfide bonds within a hypothetical protein in order to get insight into the experimental structure determination and stability of the protein. CYSPRED evaluates whether your query protein's cysteine residues form disulfide bridges/bonds. CYSPRED is a neural network-based predictor that has been taught to accurately discriminate the bonding states of cysteine in proteins, beginning with the non-binding state of the residue chain [34]. DIANNA was also employed since it aids in the prediction of disulfide connections in a protein sequence input. Understanding the function of a hypothetical protein and tertiary prediction techniques rely heavily on the ability to accurately estimate disulfide bridges [35]. We will be able to identify docking sites for hypothetical proteins based on their tertiary structure, moving one step closer to creating drugs that target diseases caused by mutations in the hypothetical gene.

Ligand binding site prediction

To anticipate protein-ligand binding sites in hypothetical proteins, the Galaxy server was employed. GalaxySite predicts the ligand-binding site of a query protein based on its tertiary structure by protein-ligand docking. The structure may be either an experimental structure (with or without ligand) or a model structure. If a protein sequence is provided, GalaxySite predicts the structure by using the GalaxyTBM method without refinement step. The binding ligands are predicted from the complex structures of similar proteins detected by HHsearch.

The protein-ligand complex structures are then predicted by a ligand docking method called LigDockCSA [36].

Detecting active sites

This protein's active site was determined by using the Computed Atlas of Surface Topography of Proteins (CASTp), located at http://sts.bioengr.uic.edu/castp/ [37]. It is a web-based tool for identifying, defining, and quantifying concave surface areas on 3D protein structures.

Studies on molecular docking and simulation

The molecular docking and simulation studies were carried out using the ClusPro v2.0 server. The algorithm running behind the ClusPro v2.0 server is very robust and does not require any prior information regarding either template or binding site between the protein-protein. The online server can be accessed at (https://cluspro.bu.edu/). ClusPro provides a user-friendly interface for understanding flexible docking between protein-protein interactions [38]. A detailed insight into the docking procedure is provided by the server whilst providing complete flexibility to the peptide sequence as well as providing permissible flexibility to the protein receptor sequence. Finally, the docking prediction result, clustering details, and interaction models generated by ClusPro were analyzed. The final docked protein-protein complex was visualized in Discovery studio. Further, the amino acid interactions occurring between protein-protein complexes were tabulated using PDBsum [39]. HawkDock was used to calculate the binding energy of a protein-protein complex [40]. To estimate interaction free energies between diverse protein-protein complexes, the service uses molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) [40].

Result and Discussions

Similarity identification, Multiple sequence alignment, and phylogeny analysis

The results of BLASTp against a non-redundant database revealed similarities with other flagellum proteins (Tables 1). The FASTA sequences of the hypothetical protein (HP) (AKH89314.2) and homologous identified proteins were aligned using multiple sequence alignment (Supplementary Fig. 1). To corroborate homology assessments of proteins at the

complex and subunit levels, phylogenetic analysis was used. The alignment and BLAST results were used to create a phylogenetic tree, which offers a comparable idea about the protein (Fig. 1). The distances between branches are also taken into consideration.

Table 1. Non-redundant sequencing yielded a protein with similar properties

Protein ID	**Accession No. AKH89314.2**			
	Organism	**Protein Name**	**Identity (%)**	**e value**
WP_200900965.1	*Edwardsiella tarda*	hypothetical protein	100	0.00
WP_207769806.1	*Edwardsiella tarda*	hypothetical protein	99.78	0.00
WP_109579113.1	*Edwardsiella tarda*	sodium:solute symporter	98.90	0.00
WP_217700654.1	*Edwardsiella tarda*	hypothetical protein	98.46	0.00
WP_024524640.1	*Edwardsiella hoshinae*	sodium:solute symporter	95.18	0.00

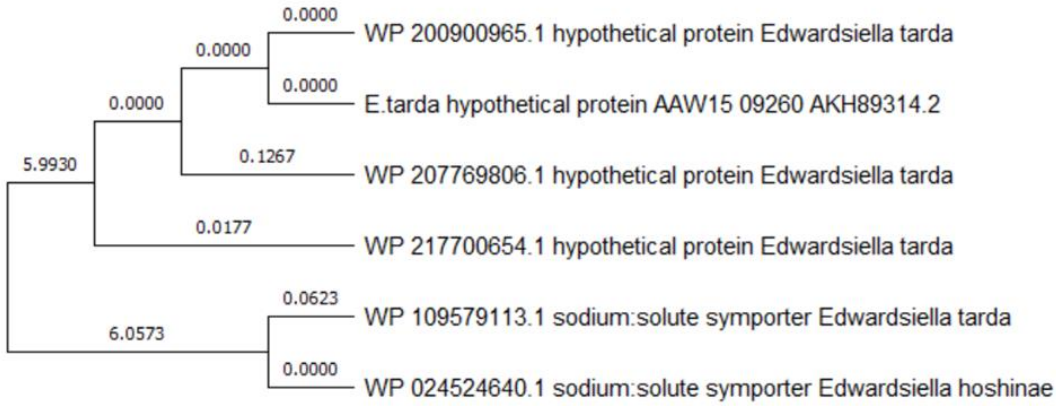


Fig 1 Multiple sequence phylogenetic tree

Physicochemical features

The protein consists of 384 amino acids, among the most abundant was Ala (A) 65 followed by, Leu (L) 58, Gly (G) 44, Ile (I) 40, Val (V) 38, Ser (S) 33, Phe (F) 26, Thr (T) 26, Arg (R) 19, Tyr (Y) 17, Pro (P) 15, Met (M) 13, Gln (Q) 13, Glu (E) 12, Trp (W) 11, Asn (N) 6, Asp (D) 6, His (H) 6, Lys (K) 5, and Lys (K) 5. The computed molecular weight was 48649.51 Da, with a theoretical pI of 9.22, indicating a positively charged protein. The total number of

positively charged (Arg + Lys) and negatively charged (Asp + Glu) residues were discovered to be 18 and 24, respectively. The protein was classified as unstable by the computed instability index of 35.49. The aliphatic index was 122.24, indicating that proteins are not stable across a wide temperature range. The GRAVY value was 0.899. GRAVY with a positive value implies that the protein is polar. Mammalian reticulocytes (in vitro) were found to have a half-life of 30 hours, yeast, > 20 hours, and Escherichia coli, > 10 hours. And the molecular formula of protein was identified as C2274H3544N560O588S16.

Table 2. Physiochemical features of hypothetical protein from different tools and server

ProtParam tool		**EMBOSS Pepstats**	
Sequence ID	AKH89314.2	Charge	9.0
Family (Pfam)	Sodium:solute symporter family	Improbability of expression in inclusion bodies	0.751
Domain (ScanProsite)	Sodium:solute symporter family	Average residue weight	106.688
Alignment	21–403	A280 extinction coefficients 1 mg ml-1	1.764
HMM length	383	A280 molar extinction coefficients	85830
Bit score	10.446	Tiny (A + C + G + S + T)	171
E value	3.3×10^{-112}	Small (A + B + C + D + G + N + P + S + T + V)	236
Number of AA	456	Aliphatic (A + I + L + V)	201
MW	48649.51 Da	Aromatic (F + H + W + Y)	60
pI	9.22	Non-polar (A + C + F + G + I + L + M + P + V + W + Y)	330
Extinction coefficients	42985	Polar (D + E + H + K + N + Q + R + S + T + Z)	126
Instability index	35.49	Charged (B + D + E + H + K + R + Z)	48
Aliphatic index	122.24	Basic (H + K + R)	30
GRAVY	0.899	Acidic (B + D + E + Z)	18

Hypothetical protein functional annotation

This potential protein sequence was discovered to have only a domain using the conserved domain search tool which is the SLC5-6-like_sbd superfamily (accession No. cl00456). Two further domain search tools, InterProScan and Pfam, were used to verify the result. Pfam

server predicted the Sodium: solute symporter family domain at 21–403 amino acid residues with an e-value of 3.3e-37. InterproScan server predicted Sodium/solute symporter domain (accession No. IPR032386) at 2–442 amino acid residues and Na/Glc_symporter_sf domain (accession No. IPR038377), at 2-448 amino acid residues. The solute is transported across the cytoplasmic membrane of prokaryotic and eukaryotic cells by sodium/substrate symport (or co-transport). Utilizing the energy stored in an electrochemical sodium gradient to drive the accumulation of solutes against a concentration gradient (sodium motive force, SMF).

Nature of subcellular localization

It is very important for predicting the subcellular localization of a hypothetical protein to further design a vaccine or drug target. It is also important to elucidate protein function. CELLO predicted subcellular localization analysis, which was confirmed by PSORTb, SOSUIGramN, and PSLpred. The HP's subcellular location was anticipated to be the inner membrane (Table 3). In contrast to THMM and HMMTOP, the inner membrane protein is highly predicted to contain transmembrane helices. The query protein was also predicted to be a transmembrane protein by the CCTOP server. All of these findings point to the protein being an inner membrane.

Table 3. Sub-cellular localization of hypothetical protein predicting from different servers

No.	Analysis	Result
1.	CELLO 2.5	InnerMembrane
2.	PSORTb	CytoplasmicMembrane
3.	SOSUIGramN	InnerMembrane
4.	PSLpred	InnerMembrane
5.	TMHMM 2.0	12 transmembrane helices present
6.	HMMTOP	25 transmembrane helices present
7.	CCTOP	Transmembrane protein

Secondary structure analysis

The secondary structures play important roles in protein structure and protein folding. All proteins functions are dependent on their structure, which, in turn, depends on physical and chemical parameters. In our study, the proportions of alpha-helix, beta-sheet content, coil

content, and overall confidence value were 79 %, 1 %, 32 percent, 20 %, and 85.2%, respectively, according to the PROTEUS Structure Prediction Server 2.0.

3D structure prediction, model quality refinement, and assessment

The three-dimensional structure of the target protein was predicted using the Raptorx server (http://raptorx.uchicago.edu/) and protein model 1 was chosen. The RaptorX program predicts 3D structures for protein sequences that have no close homologs in the Protein Data Bank (PDB) developed by the Xu group. A sequence input is used to predict secondary and tertiary structures, solvent accessibility, disordered regions, and solvent accessibility, according to RaptorX [41]. The Galaxy Refine server was used to refine the protein's projected tertiary structure, yielding five refined models and increasing the number of amino acid residues in the favored location. When compared to the other models, the scores listed above indicate the improved model's caliber. Tertiary model and refine model 1 were chosen and visualized in Discovery Studio (Figure 2 A and B). Through a Ramachandran plot analysis, PROCHECK evaluated the scalability of the galaxy server refined model, where the distribution of φ and ψ angles according to the model limits are depicted in Fig. 2C. A valid model covers 93.4% of the residues in the most preferred regions. A 3D structure model of the target sequence was validated by Verify3D and ERRAT and then compared against the established model. On the Verify3D graph, 91.77% of residues have an average 3D-1D score of ≥ 0.2, showing that the model has an excellent environmental profile, and the overall quality factor of 93.0804 in ERRAT indicates that the model is good. The YASARA energy minimization server later modified the 3D structure. Before energy minimization, the computed energy was –70,730.4 kJ/mol, but after energy minimization (by three rounds of steepest descent approach), it was reduced to –234,736.5 kJ/mol, making the modeled structure more stable. In addition, ProSA web server analysis resulted in a Z score of -7.01 which indicates the model validation (Figure 2D).

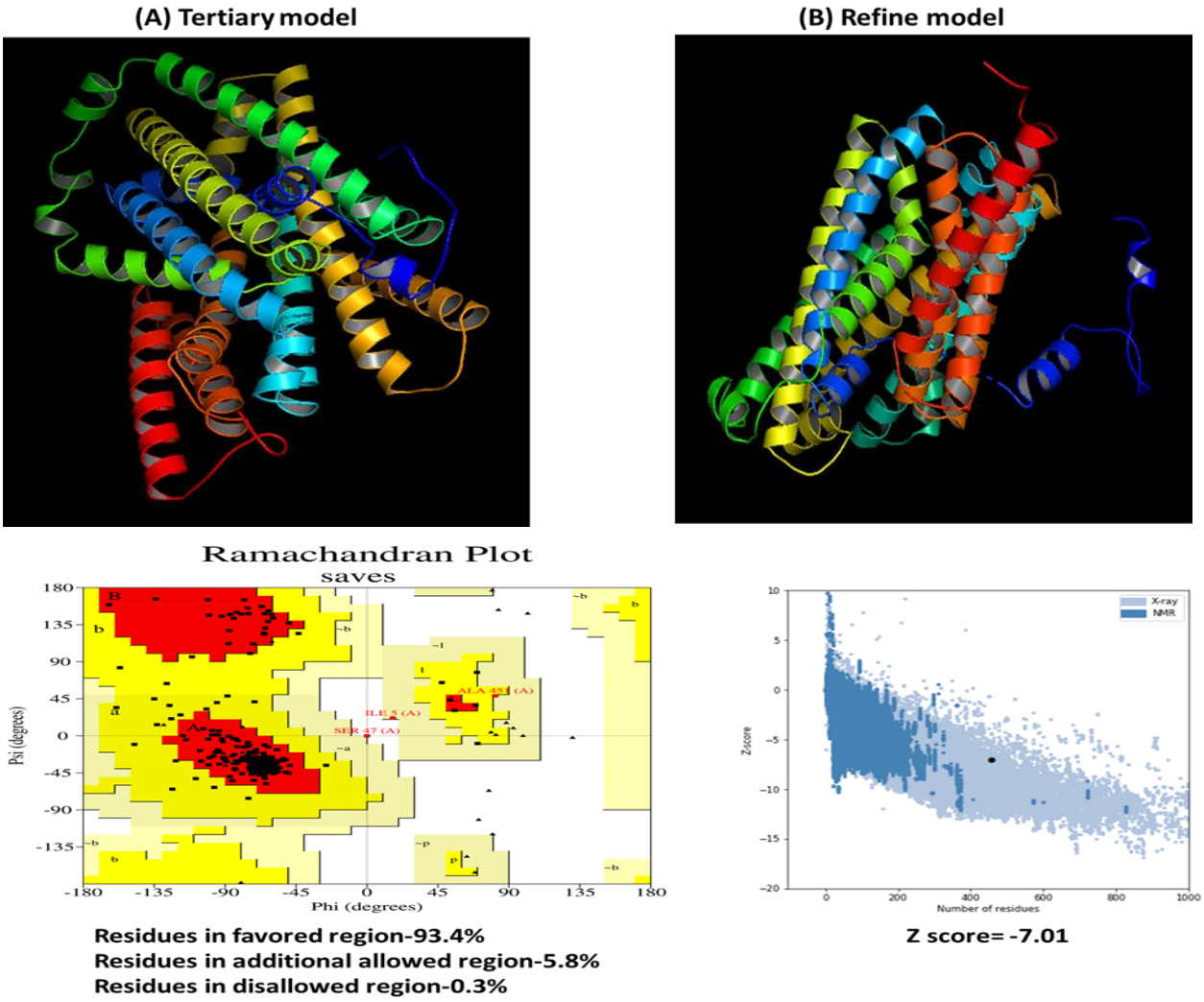


Fig 2 (A) Predicted tertiary structure of the hypothetical protein, (B) Refine model of the hypothetical protein from Galaxy refine server, (C) Ramachandran plot analysis of the refined model, and (D) Z-score results of the refined model from ProSA server.

Analysis of protein-protein interactions

We used the STRING 10.0 algorithm to make a prediction regarding the protein's possible functional interactions [31]. The identified functional partners with scores were; putA (0.905), GAC65625.1 (0.871), GAC65628.1(0.706), nanM (0.678), nanE (0.589), GAC63966.1 (0.554), nanA (0.524), GAC65383.1 (0.510) and acs (0.461). Of them, putA is a possible Proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase protein.

nanE and nanA is a Putative n-acetylmannosamine-6-phosphate 2-epimerase and N-acetylneuraminate lyase enzyme respectively (Fig 3).

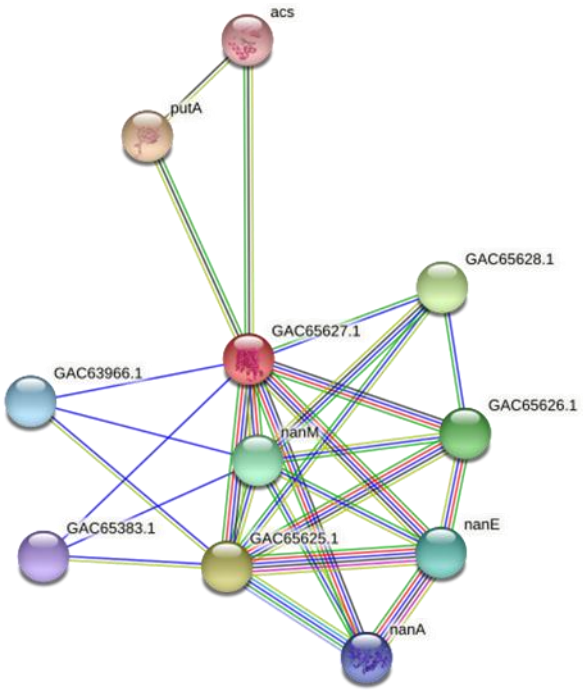


Fig 3. String (Protein-protein interactions) analysis of hypothetical protein

Table 4. CYSPRED and DIANA predict cysteine residues important in disulfide bonding

CYSPRED			**DIANA**	
Cysteine	**Prediction**	**Reliability**	**Distance**	**Bonded cysteine**
CYS 250	NON-Bonding State	9	184	AERQICQTQHY-VAQLSCAPVTA
CYS 404	NON-Bonding State	9		
CYS 435	NON-Bonding State	9		

Ligand binding interactions

Galaxy server ligand binding site predictions were done by matching target models with the PDB file of the best-predicted domain-A model. Three models were predicted by a galaxy server with different ligands. Galaxy server also combines the results into three parts Predicted ligand-binding residues, Predicted binding poses of the model, and Templates for protein-ligand complex (Table 5; Fig. 4 (A, B)). The details of the protein-ligand interaction

analysis were given in Table 5. The most probable protein-ligand binding poses and templates model for another protein-ligand complex was given in Fig 4.

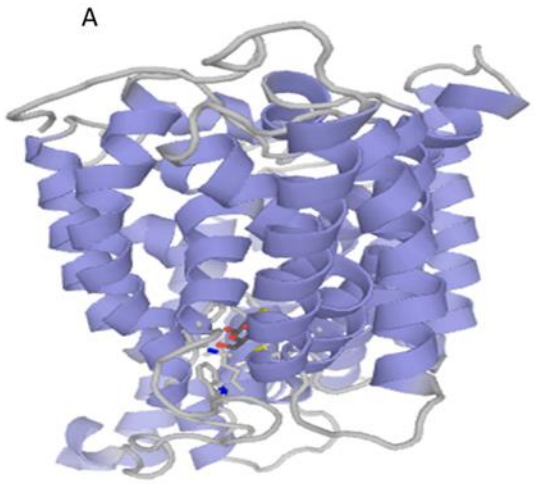
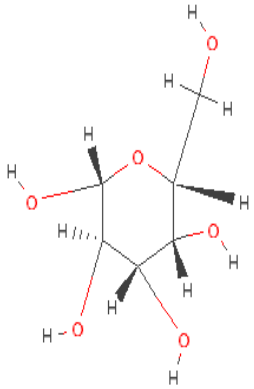
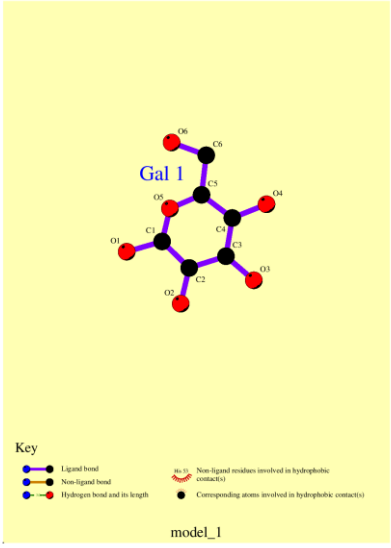
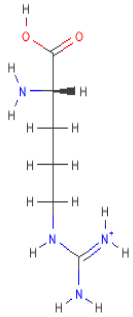
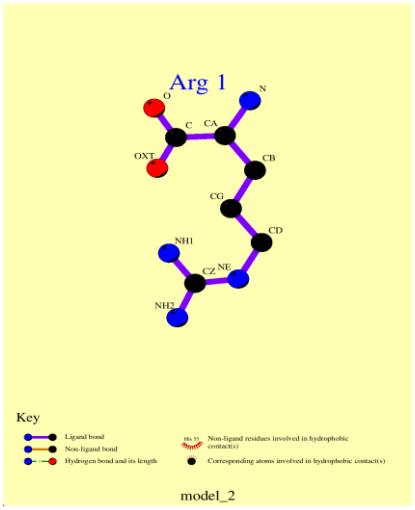


Fig. 4 (A) Predicted binding poses

Table 5. Predicted ligand-binding residues

No	Ligand Name	Ligand Structure	Binding residues	Interaction analysis
1	GAL		53K 65Y 210W 219M 364I 368M	
2	ARG		40T 42I 43G 44G 45G 122Q 168Q 169F 172I 304A 308S	

Active Site Detection

As predicted by the CASTp v.3.0 algorithm, the protein modeled contains 28 unique active sites (Fig 6). CASTp is a database server that can recognize regions on proteins, determine their boundaries, compute the area of the areas, and calculate the dimensions of the areas. Vacuums concealed within proteins and pockets on protein surfaces are also involved. To

define a pocket and volume spectrum or vacuum, surfaces of solvent-accessible molecules (Richard surface) and molecular surfaces (Connolly surface) are employed. CASTp might be utilized to look at the operational zones and surface properties of proteins. CASTp provides a dynamic, graphical user interface as well as on-the-fly measuring of user-submitted constructs [42]. Based on the area of 582.986 and the volume of 315.671, the top active sites of the model protein were identified (Fig 5). Fig 5 shows the protein's anticipated active site together with its amino acid residues.

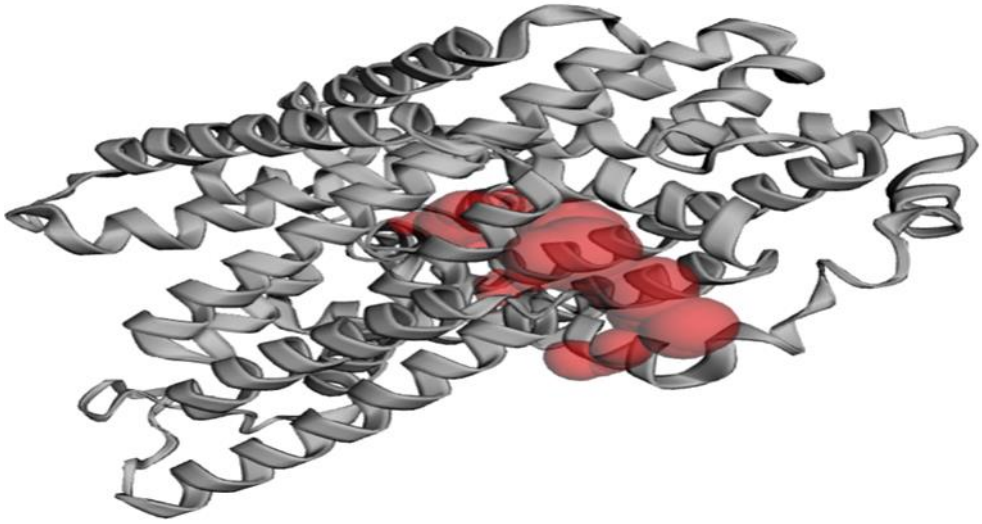


Fig 5. The hypothetical protein's active location. The red sphere represents the protein's active site

Fig 6. Active location of the hypothetical protein. The active site of amino acid residues (Blue color)

Molecular docking and simulation studies between hypothetical protein and chemotaxis protein

The docking prediction performed by the Cluspro server result showed 10 models for the docked complex. Model 1 (Figure 7 A) is considered to be the most probable model. After the selection of an appropriate model, a detailed investigation was done into the amino acid sequences that interact between the protein Chemotaxis and amino acid sequence of the hypothetical protein. The amino acid interactions revealed that the complex is stabilized by 2 hydrogen bonds (Fig 7B). Within the putA-hypothetical protein complex, Val1 and Thr91 of putA protein respectively interact with Gln192 and Leu292 of chemotaxis protein through H-bonds (Fig 7B). These H-bonds contribute towards the stability of the complex. The binding free energy of the protein complex was evaluated by the HawkDock server and found to be – 51.92 (kcal/mol) proving that the complex is stable. Additionally, three different binding sites of the protein complex was shown in Fig 7 (C, D, E).

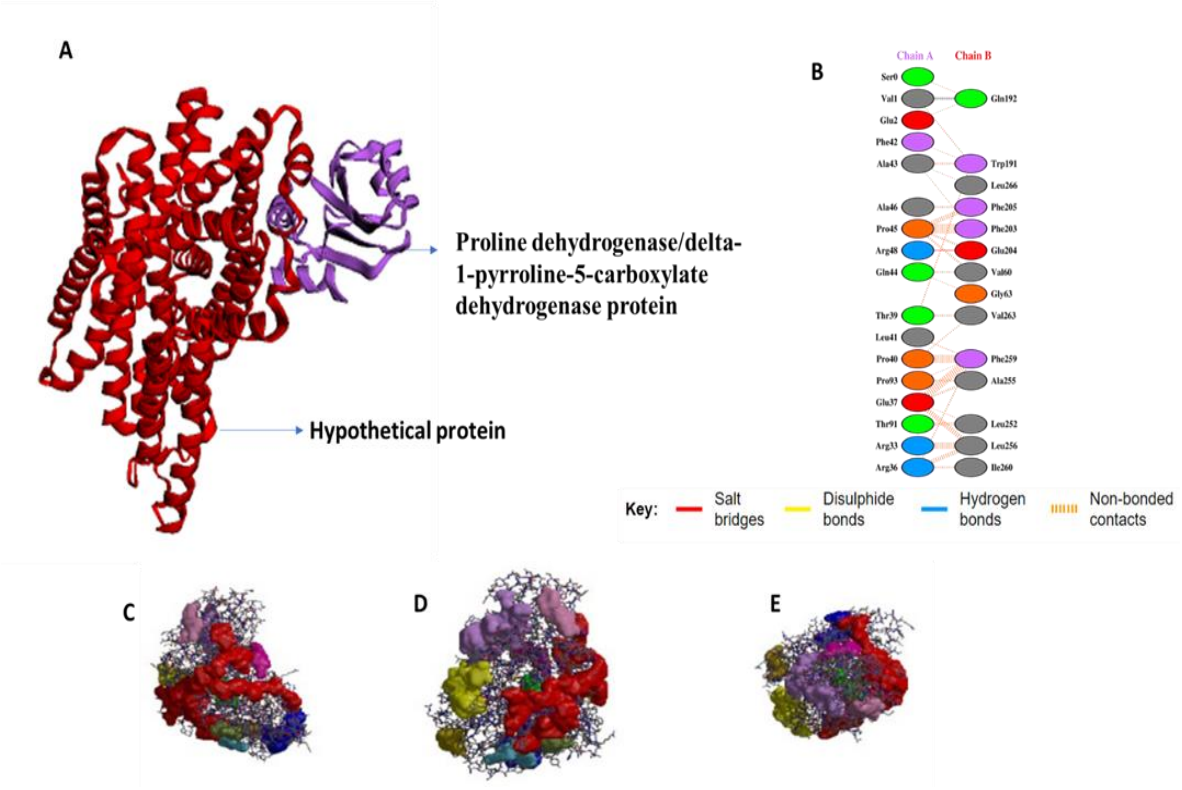


Fig. 7 (A) Chemotaxis-hypothetical protein complex, (B) Interactions between the protein-protein complex and (C, D, E) Three different binding sites of the selected protein-protein complex

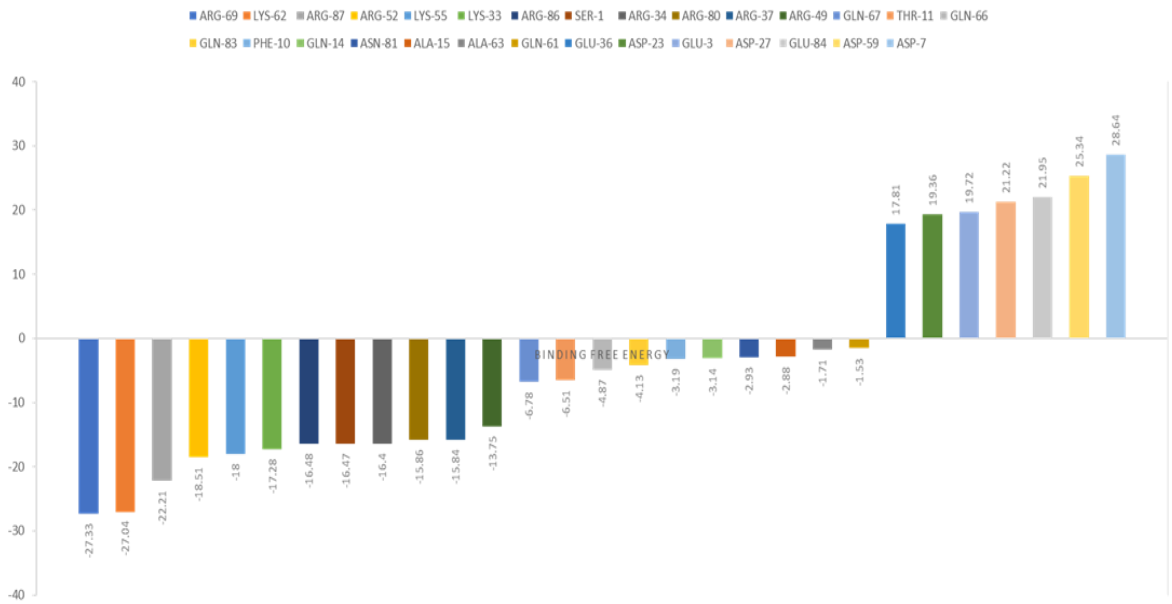


Fig 8 Bar graph of top 30 bindings free energy from HawkDock server

Conclusion

The hypothetical protein domain has a crucial role as a host cell invasion and apoptosis, according to the research. It was also discovered to be a soluble protein with a single exposed domain. The existence and distribution of this hypothetical protein domain across a wide range of bacterial strains and interactions with putA protein suggest that new antibacterial drugs could be developed. More research is being done, such as protein-ligand docking studies, to identify the representative amino acids involved in ligand binding. The molecular docking and simulation studies between hypothetical protein and putA protein are found in a stable interaction therefore it may be of interest to researchers looking to produce new drugs against multidrug resistance of *Edwardsiella tarda*.

Acknowledgments
The first author is sincerely grateful to the ASEAN and Non-ASEAN scholarship authority at Chulalongkorn University, Thailand for giving financial support for pursuing masters studies.

References

1. Hassan, H.A., et al., Fish borne Edwardsiella tarda eha involved in the bacterial biofilm formation, hemolytic activity, adhesion capability, and pathogenicity. Arch Microbiol, 2020. 202(4): p. 835-842.
2. Mohanty, B.R. and P.K. Sahoo, Edwardsiellosis in fish: a brief review. Journal of Biosciences, 2007. 32(3): p. 1331-1344.

3. Lima, L., et al., Isolation and characterizaton of Edwardsiella tarda from pacu Myleus micans. Arquivo Brasileiro De Medicina Veterinaria E Zootecnia - ARQ BRAS MED VET ZOOTEC, 2008. 60.
4. Damme, L. and J. Vandepitte, Frequent isolation of Edwardsiella tarda and Pleisiomonas shigelloides from healthy Zairese freshwater fish: A possible source of sporadic diarrhea in the tropics. Applied and environmental microbiology, 1980. 39: p. 475-9.
5. Zhang, M., J. Sun, and L. Sun, Regulation of autoinducer 2 production and luxS expression in a pathogenic Edwardsiella tarda strain. Microbiology (Reading, England), 2008. 154: p. 2060-9.
6. Sun, J., et al., Genetic Mechanisms of Multi-Antimicrobial Resistance in a Pathogenic Edwardsiella tarda Strain. Aquaculture, 2009. 289: p. 134-139.
7. Wang, Y., X.-H. Zhang, and B. Austin, Comparative analysis of the phenotypic characteristics of high- and low-virulent strains of Edwardsiella tarda. Journal of fish diseases, 2010. 33: p. 985-94.
8. Zain, N., N. SalİKan, and K. Kam, Isolation of Antibiotic Resistant Bacteria from Rivers in Terengganu, Malaysia. International Journal of Life Sciences and Biotechnology, 2020.
9. Done, H.Y., A.K. Venkatesan, and R.U. Halden, Does the Recent Growth of Aquaculture Create Antibiotic Resistance Threats Different from those Associated with Land Animal Production in Agriculture? The AAPS Journal, 2015. 17(3): p. 513-524.
10. Watts, J.E.M., et al., The Rising Tide of Antimicrobial Resistance in Aquaculture: Sources, Sinks and Solutions. Marine Drugs, 2017. 15(6): p. 158.
11. Yu, J., et al., Large antibiotic-resistance plasmid of Edwardsiella tarda contributes to virulence in fish. Microbial pathogenesis, 2012. 52: p. 259-66.
12. Roberts, M.C., Update on acquired tetracycline resistance genes. FEMS Microbiology Letters, 2005. 245(2): p. 195-203.
13. Sakai, T., et al., Identification of a 19.3-kDa protein in MRHA-positive Edwardsiella tarda: Putative fimbrial major subunit. FEMS microbiology letters, 2003. 226: p. 127-33.
14. Reichley, S.R., et al., Complete Genome Sequence of Edwardsiella tarda Isolate FL95-01, Recovered from Channel Catfish. Genome Announc, 2015. 3(3).
15. Boeckmann, B., et al., The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Research, 2003. 31(1): p. 365-370.
16. Johnson, M., et al., NCBI BLAST: a better web interface. Nucleic Acids Research, 2008. 36(suppl_2): p. W5-W9.
17. Alzohairy, A., BioEdit: An important software for molecular biology. GERF Bulletin of Biosciences, 2011. 2: p. 60-61.
18. Gasteiger, E., et al., ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research, 2003. 31(13): p. 3784-3788.
19. Yu, C. and J. Hwang. Prediction of Protein Subcellular Localizations. in 2008 Eighth International Conference on Intelligent Systems Design and Applications. 2008.
20. Yu, N.Y., et al., PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics, 2010. 26(13): p. 1608-1615.
21. Bhasin, M., A. Garg, and G.P.S. Raghava, PSLpred: prediction of subcellular localization of bacterial proteins. Bioinformatics, 2005. 21(10): p. 2522-2524.
22. Möller, S., M.D.R. Croning, and R. Apweiler, Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics, 2001. 17(7): p. 646-653.
23. Tusnády, G.E., and I. Simon, The HMMTOP transmembrane topology prediction server. Bioinformatics, 2001. 17(9): p. 849-850.
24. Dobson, L., I. Reményi, and G.E. Tusnády, CCTOP: a Consensus Constrained TOPology prediction web server. Nucleic Acids Research, 2015. 43(W1): p. W408-W412.

25. Marchler-Bauer, A., et al., CDD: a Conserved Domain Database for protein classification. Nucleic Acids Research, 2005. 33(suppl_1): p. D192-D196.
26. Kanehisa, M., et al., The KEGG databases at GenomeNet. Nucleic Acids Research, 2002. 30(1): p. 42-46.
27. Finn, R.D., Pfam: the protein families database, in Encyclopedia of Genetics, Genomics, Proteomics, and Bioinformatics. 2005.
28. Wilson, D., et al., The SUPERFAMILY database in 2007: families and functions. Nucleic Acids Research, 2006. 35(suppl_1): p. D308-D313.
29. Hunter, S., et al., InterPro: the integrative protein signature database. Nucleic Acids Research, 2008. 37(suppl_1): p. D211-D215.
30. Shen, H.-B., and K.-C. Chou, Predicting protein fold pattern with functional domain and sequential evolution information. Journal of Theoretical Biology, 2009. 256(3): p. 441-446.
31. McGuffin, L.J., K. Bryson, and D.T. Jones, The PSIPRED protein structure prediction server. Bioinformatics, 2000. 16(4): p. 404-405.
32. Xu, J., M. McPartlon, and J. Li, Improved protein structure prediction by deep learning irrespective of co-evolution information. Nat Mach Intell, 2021. 3: p. 601-609.
33. Szklarczyk, D., et al., STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res, 2015. 43(Database issue): p. D447-52.
34. Grützner, A., et al., Modulation of titin-based stiffness by disulfide bonding in the cardiac titin N2-B unique sequence. Biophysical Journal, 2009. 97(3): p. 825-834.
35. Ferrè, F. and P. Clote, DiANNA: a web server for disulfide connectivity prediction. Nucleic Acids Res, 2005. 33(Web Server issue): p. W230-2.
36. Heo, L., et al., GalaxySite: ligand-binding-site prediction by using molecular docking. Nucleic Acids Res, 2014. 42(Web Server issue): p. W210-4.
37. Dundas, J., et al., CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Research, 2006. 34(suppl_2): p. W116-W118.
38. Kozakov, D., et al., The ClusPro web server for protein-protein docking. Nature protocols, 2017. 12(2): p. 255-278.
39. Laskowski, R.A., et al., PDBsum: Structural summaries of PDB entries. Protein Sci, 2018. 27(1): p. 129-134.
40. Weng, G., et al., HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. Nucleic Acids Res, 2019. 47(W1): p. W322-w330.
41. Källberg, M., et al., RaptorX server: a resource for template-based protein structure modeling. Methods Mol Biol, 2014. 1137: p. 17-27.
42. Tian, W., et al., CASTp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Research, 2018. 46(W1): p. W363-W367.