Journal of Applied Biological Sciences
E-ISSN: 2146-0108 16(2): 191-205, 2022
DOI: https://doi.org/10.71336/jabs.1027
Research Article

ANALYSIS OF A HYPOTHETICAL PROTEIN FROM *VIBRIO HARVEYI* IDENTIFIED POSSIBLE CONNECTION WITH BIOPOLYMER METABOLISM: AN IN-SILICO APPROACH

Sk Injamamul Islam[1], Moslema Jahan Mou[2]

[1]*Chulalongkorn University, Faculty of Veterinary Science and Technology, Department of Veterinary Microbiology, Thailand*

[2]*University of Rajshahi, Faculty of Earth and Life Science, Department of Genetic Engineering and Biotechnology, Bangladesh*

**Corresponding Author:*
E-mail: 6378506331@student.chula.ac.th

(Received 14th December 2021; accepted 11th January 2022)

ABSTRACT. Because of our limited understanding of the mechanisms governing *Vibrio* species persistence and spread in the face of global warming, finding ways to control the increasing spread of pathogenic vibrios is challenging. To determine whether the persistence of *Vibrio harveyi* is associated with physiological and gene expression patterns, it is important to do research on several proteins in its genome which are classified as hypothetical proteins (HPs). As a result, the current work sought to elucidate the roles of a HP found in the genome of *V.harveyi*. To determine the structure, function and homologous model of this protein, quality bioinformatics methods were used to predict and confirm the function as well as 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 act as DNA binding site. Further, protein-protein interactions within selected HP revealed several functional partners that are essential for bacterial survival with different functional activity. In addition, molecular docking and simulation results showed stable bonding between HP and ExbD/TolR family protein which might be of significant relevance to future bacterial genetics research.

Keywords: *Hypothetical protein, molecular docking, ExbD/TolR protein, DNA, V. Harveyi.*

INTRODUCTION

The marine bacterium *Vibrio harveyi* is a heterotrophic Gram-negative luminous bacterium that inhabits temperate and tropical waters. *Vibrio* spp. can be found connected to biotic or abiotic surfaces in their natural environments, both free-living and in symbiotic or host-pathogen associations with other species, occasionally causing infections or even death of the affected animals [1, 2]. During the era of global warming (GW), there has been a gradual increase in sea surface temperature as a result of pathogenic vibrios spreading more widely [3]. The increase in cholera or/and gastroenteritis outbreaks has coincided with the above trends [4], which seem to be due to the consumption of water contaminated with pathogenic vibrios. In Asian countries, Vibrio-related diseases are more prevalent as the water temperature increases towards the equator.

There have also been reports of some highly virulent strains of *V. harveyi* able to cause mass mortality in both marine invertebrates [5] and various fishes [6] which are associated with the overall increase in the occurrence of Vibrio-associated diseases. In addition to its widespread occurrence in aquaculture farms and its effects on the seafood industry, *V. harveyi* is notorious for being an important pathogen of cultured penaeid shrimp. This is particularly true in tropical countries, where temperatures tend to be higher [7]. GW is anticipated to have even greater impacts on the ocean in the coming years, which could likely result in the spread of Vibrio-associated infections, as the forecast is not encouraging.

Nevertheless, many of this *V. harveyi* 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 structure of those HPs, their biological functions, and their interactions with other proteins.

It got simpler to attribute function to a HP utilizing numerous bioinformatic methods as 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 QFQ78833.1 of *V. harveyi*. 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 QFQ78833.1.

MATERIALS AND METHODS

Identification of similarity and retrieval of sequences

The NCBI database was used to obtain the sequence information for the hypothetical protein (HP) (QFQ78833.1). After that, the sequence was saved as a FASTA format and submitted to multiple prediction servers for in-silico analysis. A similarity search was conducted with the NCBI protein database to provide a first prediction regarding the function of the targeted HP against non-redundant [8] database to use the BLASTp tool to search proteins that may have similar characteristics to the HP [9].

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 [10]. The phylogenetic analysis was performed using an older version of 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, total number of positively charged residues (Arg + Lys), total number of negatively charged residues (Asp + Glu), aliphatic index, and grand average of hydropathicity (GRAVY) [11].

Subcellular localization analysis

CELLO anticipated subcellular localization [12]. The results were also compared to PSORTb subcellular localization predictions [13], PSLpred [14], and SOSUIGramN. TMHMM [15], HMMTOP [16], and CCTOP [17] 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)[18], for conserved domain. The Motif (Genome Net) server was used to find protein motifs [19]. The evolutionary connections of the protein were assigned using Pfam [20] and SuperFamily [21] database. For the functional analysis of the protein, the protein sequence analysis and classification software InterProScan [22] was used. The PFP-FunD SeqE server [23] 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) [24]. 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/) [25]. 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 [26]. 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 [26]. 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/) [27].

Assessment of model quality

Subsequently, the predicted three-dimensional structure was evaluated using PROCHECK and Verify3D Structure Evaluation server. PROCHECK was used to visualize the backbone dihedral angles ψ against φ of amino acid residues in the HP structure [28]. The aim of PROCHECK is to assess how normal, or conversely how unusual, the geometry of the residues in a given protein structure is, as compared with stereochemical parameters derived from well-refined, high-resolution structures [29]. Verify3D Determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigning a structural class based on its location and environment [30].

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 [31].

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. To gain insight into the experimental structure determination and stability of a hypothetical protein, CYSPRED and DIANA were used to predict disulfide bonds within 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 [32]. DIANA 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 [33]. 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

In order 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 [34].

Detecting active sites

The active site of this protein was determined using Computed Atlas of Surface Topography of Proteins (CASTp), available at http://sts.bioengr.uic.edu/castp/ [35]. 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 ClusPro v2.0 server. The algorithm running behind 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 [36]. A detailed insight into 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 complex were tabulated using PDBsum [37]. HawkDock server (http://cadd.zju.edu.cn/hawkdock/) was used to calculate the binding energy of a protein-protein complex [38]. To estimate interaction free energies between diverse protein-protein complexes, the service uses molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) [38].

RESULTS AND DISCUSSION

Similarity identification, Multiple sequence alignment and phylogeny analysis

The results of BLASTp against a non-redundant database revealed similarity with other flagellum proteins (Table 1). The FASTA sequences of the hypothetical protein (HP) (QFQ78833.1) 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. QFQ78833.1**	
	Organism	**Protein Name**	**Identity (%)**	**e value**
WP_005430084.1	*Vibrio campbellii*	hypothetical protein	100	0.00
WP_017817185.1	*Vibrio harveyi*	hypothetical protein	99.75	0.00
WP_045456090.1	*Vibrio campbellii*	hypothetical protein	99.75	0.00
WP_041057284.1	*Vibrio owensii*	hypothetical protein	99.26	0.00
WP_175546361.1	*Vibrio vulnificus*	hypothetical protein	99.26	0.00

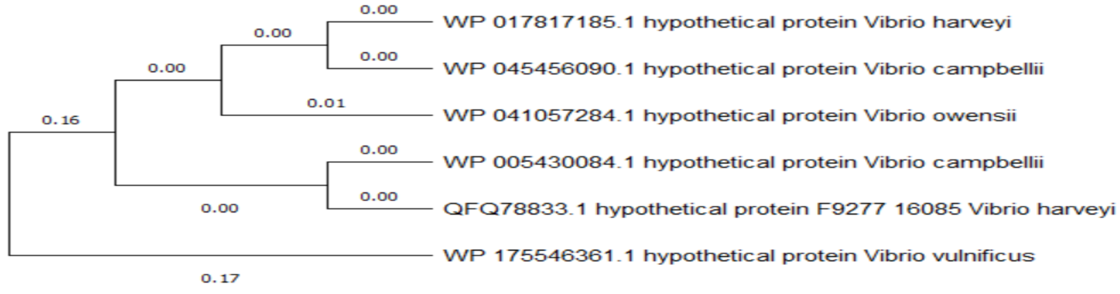


Fig 1. *Multiple sequence phylogenetic tree*

Physicochemical features

The protein consists of 404 amino acids, among the most abundant was Ala 51 followed by, Leu 45, Lys 42, Ser 31, Gln 31, Glu 27, Tyr 24, Thr 21, Val 20, Met 18, Asp 14, Ile 14, Pro 13, Arg 13, Asn 13, Phe 11, Gly 9, Trp 5 and His 2. The computed molecular weight was 45969.20 Da, with a theoretical pI of 9.39, indicating a positively charged protein. The total number of positively charged (Arg + Lys) and negatively charged (Asp + Glu) residues were discovered to be 55 and 41, respectively. The protein was classified as unstable by the computed instability index of 38.26. The aliphatic index was 83.94, indicating that proteins are stable across a wide temperature range. The GRAVY value was -0.374. GRAVY with a negative value implies that the protein is nonpolar. 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 $C_{2067}H_{3292}N_{538}O_{607}S_{18}$.

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

ProtParam tool		**EMBOSS Pepstats**	
Sequence ID	QFQ78833.1	Charge	15.0
Family (Pfam)	No hit	Improbability of expression in inclusion bodies	0.661
Domain (ScanProsite)	Tetratrico peptide repeat profile	Average residue weight	113.785
Hits (ScanProsite)	8	A280 extinction coefficients 1 mg ml-1	1.376
Length/Max. Score	71-321; 18.423	A280 molar extinction coefficients	63260
Bit score	10.446	Tiny (A + C + G + S + T)	112
E value	3.3×10^{-112}	Small (A + B + C + D + G + N + P + S + T + V)	172
Number of AA	404	Aliphatic (A + I + L + V)	130
MW	45969.20 Da	Aromatic (F + H + W + Y)	42
pI	9.39	Non-polar (A + C + F + G + I + L + M + P + V + W + Y)	210
Extinction coefficients	42985	Polar (D + E + H + K + N + Q + R + S + T + Z)	194
Instability index	38.26	Charged (B + D + E + H + K + R + Z)	98
Aliphatic index	83.94	Basic (H + K + R)	57
GRAVY	-0.374	Acidic (B + D + E + Z)	41

Hypothetical protein functional annotation

This potential protein sequence was discovered to have only a domain using the conserved domain search tool which is PEP_TPR_lipo super family (accession No. cl37187). The PEP-CTERM/exosortase system occurs strictly in Gram-negative bacterial species with the LPXTG/sortase system common to Gram-positive bacteria, analogous to the PEP-CTERM/exosortase system. The occurrence of this protein is dependent upon the presence of DNA-binding response regulator (TIGR02915) and transmembrane histidine kinase (TIGR02916) in a species. As many of the PEP-CTERM proteins in these genomes are preceded by a proposed recognition site for response regulators, the presence of tetratricopeptide repeats (TPR) suggests protein-protein interactions may be important for regulating the expression of those proteins. Two further domain search tools, InterProScan and Pfam, were used to verify the result but both servers predicted no domain family.

Nature of subcellular localization

CELLO predicted subcellular localization analysis, which was confirmed by PSORTb, SOSUIGramN, and PSLpred. The HP’s subcellular location was anticipated to be Periplasmic (Table 3). In contrast to THMM and HMMTOP, the periplasmic protein is predicted nothing to contain transmembrane helices. The query protein was also predicted not to be a transmembrane protein by the CCTOP server. All of these findings point to the protein being Periplasmic.

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

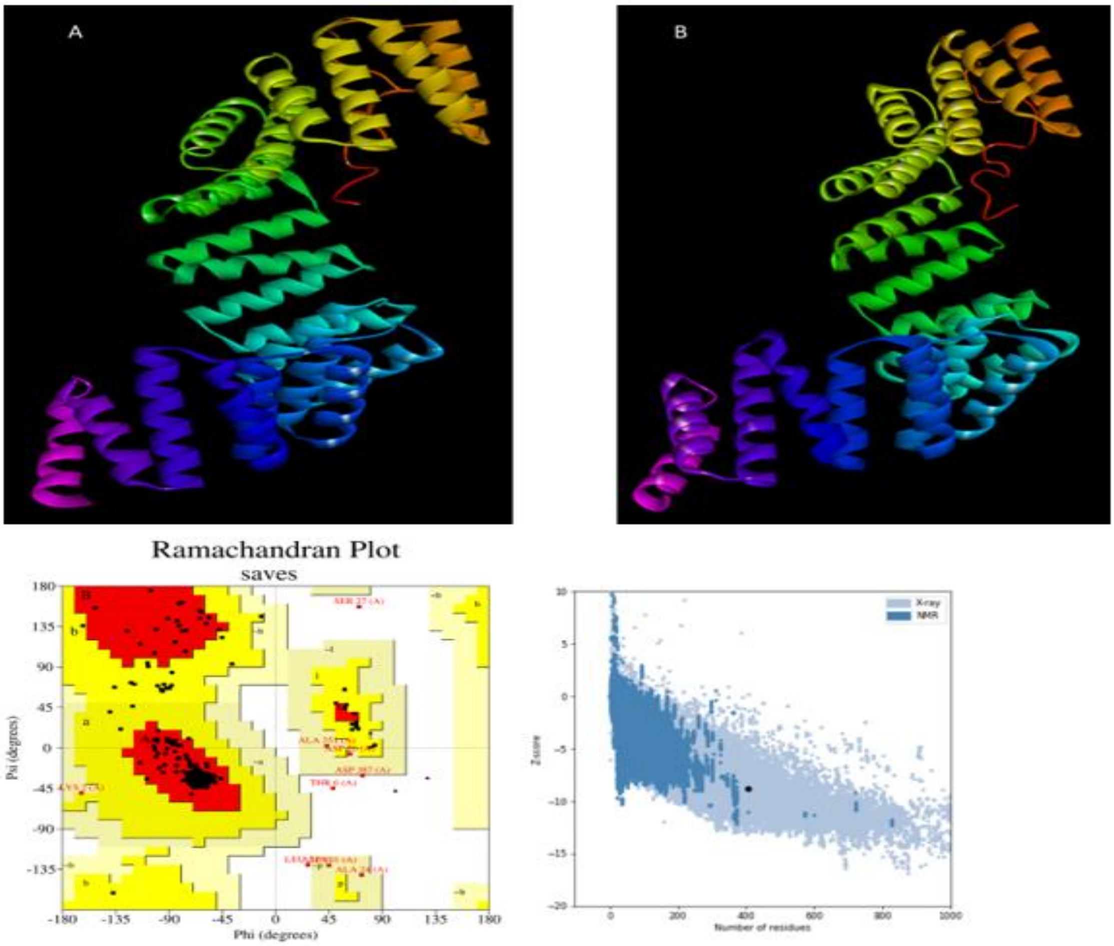
No.	**Analysis**	**Result**
1.	CELLO 2.5	Periplasmic
2.	PSORTb	Periplasmic
3.	PSLpred	Periplasmic Protein
4.	SOSUIGramN	Peripalsmic
5.	TMHMM 2.0	No transmembrane helices present
6.	HMMTOP	One transmembrane helices present (6-23)
7.	CCTOP	No Transmembrane protein

Secondary structure analysis

The secondary structures play important roles in protein structure and protein folding. PROTEUS is a web server designed to support comprehensive protein structure prediction and structure-based annotation. The PROTEUS Structure Prediction Server 2.0 predicted the proportions of alpha helix, beta sheet content, coil content, and overall confidence value were 78%, 2%, 31 percent, 19%, and 84.2%, respectively of the hypothetical protein.

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. RaptorX predicts secondary and tertiary structures, solvent accessibility, disordered regions, and solvent accessibility based on sequence input [39]. The Galaxy Refine server was used to refine the protein's projected tertiary structure, yielding five refined models and increasing the amount 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 2A 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 89.5 % 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, 92.57% 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 99.2405 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 –71,120.4 kJ/mol, but after energy minimization (by three rounds of steepest descent approach), it was reduced to –224,539.5 kJ/mol, making the modeled structure more stable. In addition, ProSA web server analysis resulted Z score of -8.83 which indicate the model validation (Figure 2D).



Residues in most favored regions: 89.5%
Residues in additional allowed regions: 8.1%
Residues in disallowed regions: 1.0%

Z Score: -8.83

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 refine model and (D) Z-score results of the refine 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; Biopolymer transport ExbD/TolR family protein (0.873), TonB family C-terminal domain protein (0.836), DUF3450 domain-containing protein (0.777), MotA/TolQ/ExbB proton channel family protein (0.756), nanE (0.589), MotA/TolQ/ExbB proton channel family protein (0.693), Uncharacterized protein (0.636), DUF3450 domain-containing protein (0.609), Periplasmic protein TonB (0.574), MotA/TolQ/ExbB proton channel family protein (.491) and Extracellular lipase, Pla-1/cef family protein (0.486). Of them, biopolymer transport ExbD/TolR family protein (AMF98084.1) is a naturally occuring biopolymers which can enable vibrios to successfully strive in marine ecosystems (fig-3).

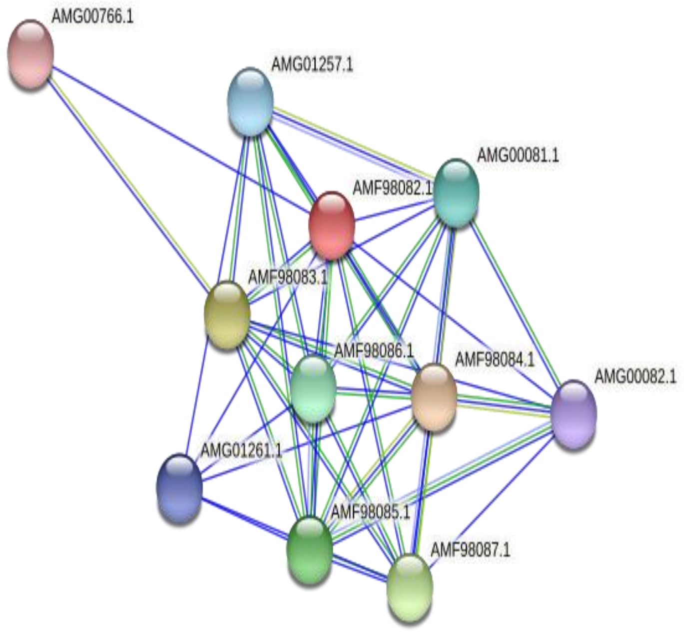


Fig 3. *String (Protein-protein interactions) analysis of hypothetical protein*

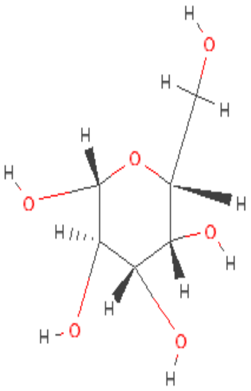
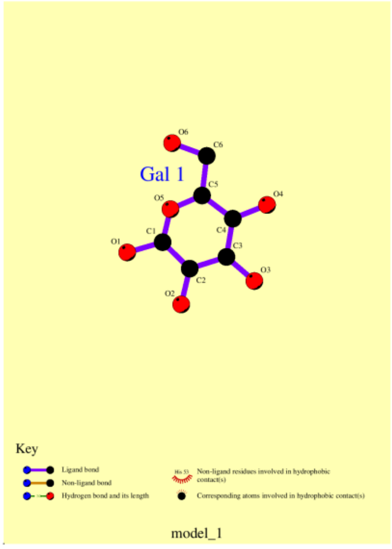
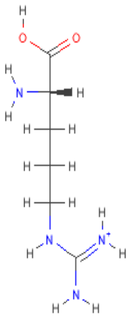
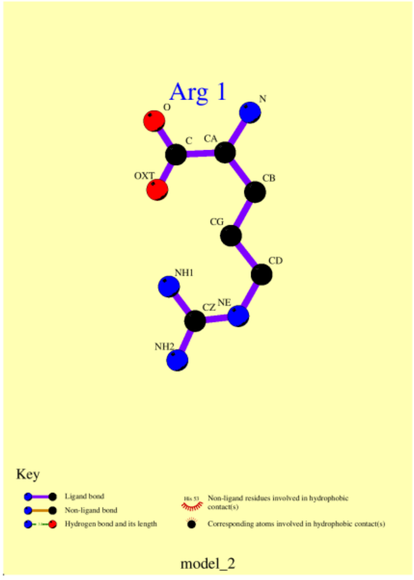
Table 4. *CYSPRED and DIANA predict cysteine residues important in disulphide bonding*

	CYSPRED		**DIANA**	
Cysteine	**Prediction**	**Reliability**	**Distance**	**Bonded cysteine**
CYS 12	Bonding State	8	56	QEWKECPDYVS-AGENSCYFNSS

Ligand binding interactions

To predict the binding sites of ligands on Galaxy servers, we compared target models and the PDB file of the best predicted domain A model. Different ligands were predicted by the galaxy server for three models. Galaxy server also combines the results into three parts such as Predicted ligand-binding residues, Predicted binding poses of model and Templates for protein-ligand complex (Table 5). The details of the protein-ligand interaction analysis were given in Table 5.

***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 39 unique active sites (Figure 5). 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 [40]. Based on the area of 1200.598 and the volume of 3600.556, the top active sites of the model protein were identified (Figure 4). Figure 4 shows the protein's anticipated active site together with its amino acid residues.

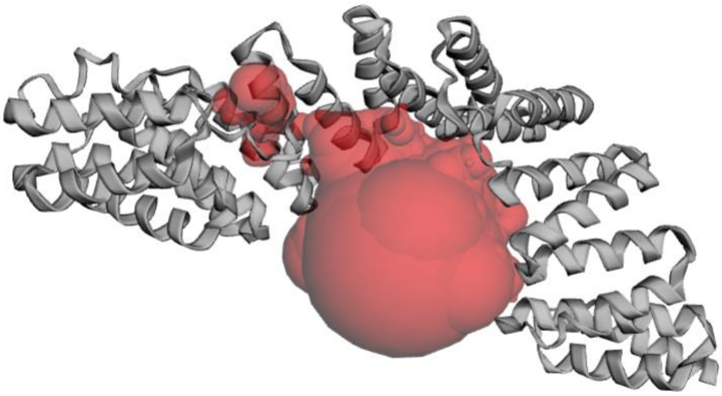


***Fig 4**. The hypothetical protein's active location. The red sphere represents the protein's active site*

Chain A

M K L L K T L L A S V A L T M G V L S T P V V A K D S P Q L S D K T F K V V N K V Q E L I A T E K Y S D A I
E R L N K A L G K T S K K Y D R A V L L Q Q M G F L Y S M R D D Y V K A S K Y F A E A L S L D A L P V P V A
Q Q V R Y S L A Q L Y L A E E Q F K K S V K T M E K W F A V A E T T K E K P Q A H A Y I T L A S A Y V Q M E
D Y R K A I A P T K K A I A M T K N P S E S W Y M L L M A S H Y E L G Q M K S V A S V L K I L T T K Y P Q K
K R Y W M Q L S G T Y M E L N Q D K N A L S A L E M A Y K L G L F E E E K E Y L R L V N F L A Y Q E I P Y R
A A K I L K D E M A N G N I E T N Q E N L E R L A S F W Q Q S K E L D N A I D T Y Q T A Y Q L A P K T K S Q
L K I A R L M I Q A K Q Y Q A A S K F A A T P A S D A S R S Q A A E L Q Y I R G M A F F E L N Q P T Q A L K
S M K K A A Q S D S M R Q V V S P W I N Y L E S Q G

***Figure 5**. 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 Cluspro server result showed 10 models for the docked complex. Model 1 (Figure 6 A) is considered to be the most probable model. After 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 hypothetical protein. The amino acid interactions revealed that the complex is stabilized by 4 hydrogen bonds (Figure 6B). Within the ExbD/TolR -hypothetical protein complex, Ala 14, Glu 12, Met 20 and Ser 36 of ExbD/TolR protein respectively interact with His 149, Trp 185, Ser 19 and Lys 49 of hypothetical protein through H-bonds (Figure 6B). These H-bonds contribute towards the stability of the complex. The binding free energy of protein complex was evaluated by HawkDock server and found to be – 50.82 (kcal/mol) proving that the complex is stable. Additionally, three different binding sites of the protein complex was shown in Figure 6 (C, D, E).

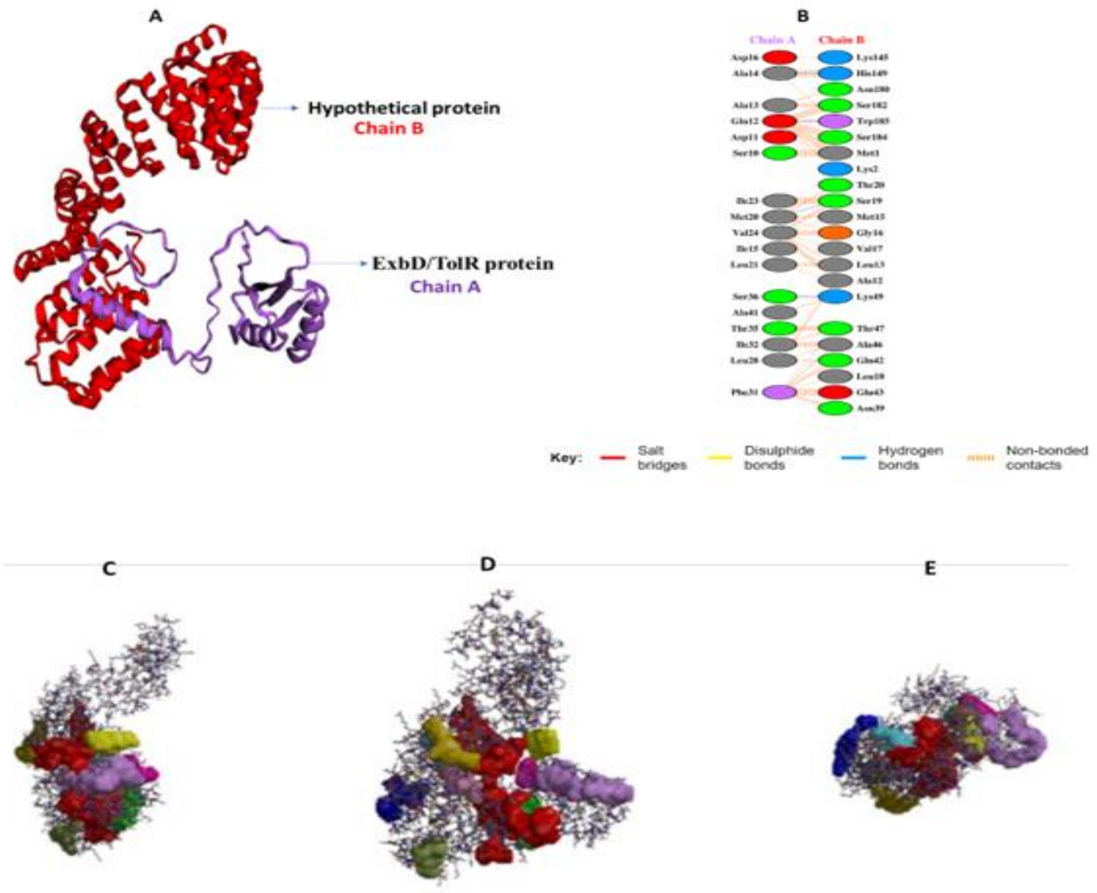


Figure 6. *(A) ExbD/TolR protein -hypothetical protein complex, (B) Interactions between the protein-protein complex and (C, D, E) Three different binding sites of the selected protein-protein complex.*

CONCLUSION

The hypothetical protein domain has a crucial role as a biopolymer metabolism, according to the research. It was also discovered to be a soluble protein with single exposed domain. The existence and distribution of this hypothetical protein domain across a wide range of bacterial strain and interaction with, ExbD/TolR protein suggests that new antibacterial drugs as well as putative mechanisms of its persistence could be revealed. More research is being done, such as protein-ligand docking studies, to identify the representative amino acids involved for ligand binding. The molecular docking and simulation studies between hypothetical protein and ExbD/TolR protein is found in a stable interaction therefore it may be of interest to researchers to know more about the biopolymer mechanisms of *V. harveyi* in the marine environment.

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

Conflict of Interest. The authors declared that there is no conflict of interest.

Authorship Contributions. Concept: S.I.I, Design: M.J.M., Data Collection or Processing: S.I.I., M.J.M., Analysis or Interpretation: S.I.I., M.J.M., Literature Search: S.I.I, Writing: S.I.I., M.J.M.

Financial Disclosure. This research received no grant from any funding agency/sector.

REFERENCES

[1] Campbell, S., Harada, R.M., DeFelice, S.V., Bienfang, P.K., Li, Q.X. (2009): Bacterial production of tetrodotoxin in the pufferfish Arothron hispidus. Natural Product Research 23(17):1630-1640.

[2] Ruwandeepika, H.A.D., Defoirdt, T., Bhowmick P.P., Shekar, M., Bossier, P., Karunasagar, I. (2010): Presence of typical and atypical virulence genes in vibrio isolates belonging to the Harveyi clade. Journal of Applied Microbiology 109(3):888-899.

[3] Roux, F.L., Wegner, K.M., Baker-Austin, C., Vezzulli, L., Osorio, C.R., Amaro, C., Ritchie, J.M., Defoirdt, T., Destoumieux-Garzón, D., Blokesch, M. (2015): The emergence of Vibrio pathogens in Europe: ecology, evolution, and pathogenesis. Frontiers in Microbiology 6(830).

[4] Baker-Austin, C., Oliver, J.D., Alam, M., Ali, A., Waldor, M.K., Qadri, F., Martinez-Urtaza, J. (2018): *Vibrio* spp. infections. Nature Reviews Disease Primers 4(1):1-19.

[5] Martin, G., Rubin, N., Swanson, E. (2004): Vibrio parahaemolyticus and V. harveyi cause detachment of the epithelium from the midgut trunk of the penaeid shrimp Sicyonia ingentis. Diseases of aquatic organisms 60:21-29.

[6] Lee, K.K., Liu, P-C., Chuang, W.H. (2002): Pathogenesis of Gastroenteritis Caused by Vibrio carchariae in Cultured Marine Fish. Marine biotechnology (New York, NY) 4:267-277.

[7] Austin, B., Zhang, X.-H. (2006): Vibrio harveyi: a significant pathogen of marine vertebrates and invertebrates. Letters in Applied Microbiology 43(2):119-124.

[8] Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I. (2003): The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Research 31(1):365-370.

[9] Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., Madden, T.L. (2008): NCBI BLAST: a better web interface. Nucleic Acids Research 36(suppl_2): W5-W9.

[10] Alzohairy, A. (2011): BioEdit: An important software for molecular biology. GERF Bulletin of Biosciences 2:60-61.

[11] Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A. (2003): ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research 31(13):3784-3788.

[12] Islam, S., Mou, M., Sanjida, S., Mahfuj, MsE. (2022): An In-silico Approach for Identifying Phytochemical Inhibitors Against Nervous Necrosis Virus (NNV) in Asian Sea Bass by Targeting Capsid Protein. Genetics of Aquatic Organisms 6:487.

[13] Yu, C., Hwang, J. (2008): Prediction of Protein Subcellular Localizations. In: 2008 Eighth International Conference on Intelligent Systems Design and Applications: 26-28 Nov 2008. 165-170.

[14] Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J. (2010): PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26(13):1608-1615.

[15] Bhasin, M., Garg, A., Raghava, G.P.S. (2005): PSLpred: prediction of subcellular localization of bacterial proteins. Bioinformatics 21(10):2522-2524.

[16] Möller, S., Croning, M.D.R.., Apweiler, R. (2001): Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17(7):646-653.

[17] Tusnády, G.E., Simon, I. (2001): The HMMTOP transmembrane topology prediction server. Bioinformatics 17(9):849-850.

[18] Dobson, L., Reményi, I., Tusnády, G.E. (2015): CCTOP: a Consensus Constrained TOPology prediction web server. Nucleic Acids Research 43(W1): W408-W412.

[19] Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He S., Hurwitz, D.I., Jackson, J.D., Ke, Z. (2005): CDD: a Conserved Domain Database for protein classification. Nucleic Acids Research 33(suppl_1): D192-D196.

[20] Kanehisa, M., Goto, S., Kawashima, S., Nakaya, A. (2002): The KEGG databases at GenomeNet. Nucleic Acids Research 30(1):42-46.

[21] Finn, R.D. (2005): Pfam: the protein families database. In: Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics.

[22] Wilson, D., Madera, M., Vogel, C., Chothia, C., Gough J. (2006): The SUPERFAMILY database in 2007: families and functions. Nucleic Acids Research 35(suppl_1):D308-D313.

[23] Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L. (2008): InterPro: the integrative protein signature database. Nucleic Acids Research 37(suppl_1):D211-D215.

[24] Shen, H.-B., Chou, K.-C. (2009): Predicting protein fold pattern with functional domain and sequential evolution information. Journal of Theoretical Biology 256(3):441-446.

[25] McGuffin, L.J., Bryson, K., Jones, D.T. (2000): The PSIPRED protein structure prediction server. Bioinformatics 16(4):404-405.

[26] Islam S., Mou, M., Sanjida, S., Mahfuj, MsE. (2022): Functional Annotation and Characterization of a Hypothetical Protein from Pseudoalteromonas spp. Identify Potential Biomarker: An In-silico Approach. Aquatic Food Studies 2:57.

[27] Xu, J., McPartlon, M., Li, J. (2021): Improved protein structure prediction by deep learning irrespective of co-evolution information. Nat Mach Intell 3:601-609.

[28] Islam, S., Sanjida, S., Mou, M., Mahfuj, MsE., Nasir, S. (2022): 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 5:78-96.

[29] Wiederstein, M., Sippl, M.J. (2007): ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research 35(Web Server issue): W407-W410.

[30] Anderson, R., Deng, Z., Campbell, R., Jiang, X. (2005): Main-chain conformational tendencies of amino acids. Proteins 60:679-689.

[31] Colovos, C., Yeates, T.O. (1993): Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2(9):1511-1519.

[32] Laskowski, R.A., MacArthur, M.W., Thornton, J.M. (2006): PROCHECK: validation of protein-structure coordinates. In: International Tables for Crystallography 684-687.

[33] Lüthy, R., Bowie, J.U., Eisenberg, D. (1992): Assessment of protein models with three-dimensional profiles. Nature 356(6364):83-85.

[34] Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P. (2015): STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43(Database issue): D447-452.

[35] Islam, S., Mou, M., Sanjida, S., Mahfuj, MsE., Alam, M., Juthy, Y. (2022): An In-silico analysis of the molecular interactions between PmCBP-VP24 and PmCBP-VP28 protein complex to understand the initial initiating events of shrimp WSSV infection. International Journal of Life Sciences and Biotechnology

[36] Grützner, A., Garcia-Manyes, S., Kötter, S., Badilla, C.L., Fernandez, J.M., Linke, W.A. (2009): Modulation of titin-based stiffness by disulfide bonding in the cardiac titin N2-B unique sequence. Biophysical journal 97(3):825-834.

[37] Ferrè, F., Clote, P. (2005): DiANNA: a web server for disulfide connectivity prediction. Nucleic Acids Res 33(Web Server issue):W230-232.

[38] Heo, L., Shin, W.H., Lee, M.S., Seok, C. (2014): GalaxySite: ligand-binding-site prediction by using molecular docking. Nucleic Acids Res 42(Web Server issue): W210-214.

[39] Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., Liang, J. (2006): CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Research 34(suppl_2): W116-W118.

[40] Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D., Vajda, S. (2017): The ClusPro web server for protein-protein docking. Nature protocols 12(2):255-278.

[41] Laskowski, R.A., Jabłońska, J., Pravda, L., Vařeková, R.S., Thornton, J.M. (2018): PDBsum: Structural summaries of PDB entries. Protein Sci 27(1):129-134.

[42] Weng, G., Wang, E., Wang, Z., Liu, H., Zhu, F., Li, D., Hou, T. (2019): HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. Nucleic Acids Res 47(W1): W322-w330.

[43] Källberg, M., Margaryan, G., Wang, S., Ma, J., Xu, J. (2014): RaptorX server: a resource for template-based protein structure modeling. Methods Mol Biol 1137:17-27.

[44] Tian, W., Chen, C., Lei, X., Zhao, J., Liang, J. (2018): CASTp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Research 46(W1): W363-W367.