

# Functional Annotation and Characterization of a Hypothetical Protein from *Pseudoalteromonas* spp. Identify Potential Biomarker: An In-silico Approach

Sk Injamamul Islam<sup>1,\*</sup> , Moslema Jahan Mou<sup>2</sup> , Saloa Sanjida<sup>3</sup> , Sarower Mahfuj<sup>1</sup> 

<sup>1</sup>Chulalongkorn University, Faculty of Veterinary Science and Technology, Department of Veterinary Microbiology, Bangkok-10330, Thailand.

<sup>2</sup>University of Rajshahi, Faculty of Life and Earth Science, Department of Genetic Engineering & Biotechnology, Rajshahi-6205, Bangladesh.

<sup>3</sup>Jashore University of Science and Technology, Faculty of Applied Science and Technology, Department of Environmental Science and Technology, Jashore-7408, Bangladesh.

## How to cite

Islam, I.S., Mou, M.J., Sanjida, S., Mahfuj, S. (2022). Functional Annotation and Characterization of a Hypothetical Protein from *Pseudoalteromonas* spp. Identify Potential Biomarker: An In-silico Approach. *Aquatic Food Studies*, 2(1), AFS57. <https://doi.org/10.4194/AFS57>

## Article History

Received 28 November 2021

Accepted 01 March 2022

First Online 03 March 2022

## Corresponding Author

Tel.: +0646904801

E-mail:

6378506331@student.chula.ac.th

## Keywords

Hypothetical protein

Molecular docking

Biomarker

*Pseudoalteromonas* spp.

## Abstract

A wide variety of *Pseudoalteromonas* spp. found in *Kappaphycus alvarezii* (marine seaweed) causing deadliest Ice-ice disease and in *Laminaria japonica* (brown seaweed) found to cause red spots disease. However, very little is known about this pathogen and its genome characteristics. 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 *Pseudoalteromonas* spp. To determine the structure and function of this protein, many bioinformatics methods were used. 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 exopolysaccharide biosynthesis and can be a potential biomarker. Further, protein-protein interactions within selected HP revealed several functional partners that are essential for bacterial survival. In addition, molecular docking and simulation results showed stable bonding between HP and HSP90. Finally, the current work shows that the annotated HP is associated with possible protein sorting signals in the environment as well as having a stable binding with the HSP90, which might be of significant relevance to future bacterial genetics research.

## Introduction

In Asia, seaweed has been used as a food for centuries and is becoming increasingly valuable for dietary supplements, animal feed, chemicals, and biofuels. These aquaculture crops are suffering from an increase in disease and pest prevalence in recent years, which has resulted in a reduction in both quantity and value (Ward et al., 2019). Much of this production is centered in Asia, where *Pyropia*, *Undaria*, and *Saccharina* spp. are cultivated for human consumption

(Kim et al., 2017), and *Kappaphycus* and *Eucheuma* spp. are cultivated for carrageenan agar production (Lim et al., 2017). However, a common problem in the production of seaweed is diseases hampering the global seaweed economy. The decline was attributed to the intensification of aquaculture activity, which has led to an increase in diseases and pests (Tsiresy et al., 2016). Understanding diseases and pests become more important as the seaweed aquaculture industry expands. As with other aquaculture sectors including finfish and shellfish, intensive cultivation of macroalgae

leads to more frequent and damaging disease outbreaks (Shinn et al., 2015).

The genus *Pseudoalteromonas* is responsible for some serious diseases, such as the Ice-ice and red spot diseases in seaweed (Syafitri et al., 2017; Ward et al., 2019). Bacteria are known to be the main component of more diverse seaweed epibionts, and they may have played a role in the disease's development (Largo et al., 1995). In recent years, many seaweed farms ceased to exist because of the "ice-ice" problem particularly in south east Asia (Li et al., 2022). Recent reports have linked disease in seaweeds, to opportunistic microbial pathogens and in some instances the causative agent (s) have been identified and virulence traits described (Li et al., 2022). However, up to now no effective strategies are available for disease control in sea weeds. 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 became easier to attribute function to an HP using various bioinformatic 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 multispecies hypothetical protein (accession no. WP\_058430053.1) of *Pseudoalteromonas spp.* 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 WP\_058430053.1.

## Methods

### Sequence Retrieval and Similarity Identification

The NCBI database was used to obtain the sequence information for the hypothetical protein (HP) (WP\_058430053.1). 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 (<https://www.ncbi.nlm.nih.gov/>) against non-redundant (Boeckmann et al., 2003) database using the BLASTp tool (<https://blast.ncbi.nlm.nih.gov/Blast>) to search proteins that may have similar characteristics to the HP (Johnson et al., 2008).

### Phylogenetics Analysis

Multiple sequence alignments were performed using the BioEdit version 7.0.5.3 biological sequence

alignment editor between the HP and proteins with similar structural characteristics to the HP (Alzohairy, 2011). The phylogenetic analysis was performed using the MEGAX 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, a 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) (Gasteiger et al., 2003).

### Subcellular Localization Analysis

CELLO anticipated subcellular localization (Yu & Hwang, 2008). The results were also compared to PSORTb subcellular localization predictions (Yu et al., 2010), PSLpred (Bhasin et al., 2005), and SOSUIGramN. TMHMM (Möller et al., 2001), HMMTOP (Tusnády & Simon, 2001), and CCTOP (Dobson et al., 2015) 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) (Marchler-Bauer et al., 2005), for the conserved domain. The Motif (Genome Net) server was used to find protein motifs (Kanehisa et al., 2002). The evolutionary connections of the protein were assigned using Pfam (Finn, 2005) and SuperFamily (Wilson et al., 2006) database. For the functional analysis of the protein, the protein sequence analysis, and classification software InterProScan (Hunter et al., 2008) were used. The PFP-FunD SeqE server (Shen & Chou, 2009) was used to recognize protein folding patterns.

### Prediction of Secondary Structure

PSI-blast based secondary structure Prediction (PSIPRED) (McGuffin et al., 2000) was applied for the prediction of the proteins' secondary structure by using a primary sequence of the protein with default parameters. A technique for studying protein structure is called PSIPRED (PSI-blast-based secondary structure PREDiction). 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 helices, and coils) based on its primary sequence.

### Three-dimensional Structure Prediction, Refinement, and Validation

The three-dimensional structure of the target protein was predicted using the RaptorX server (<http://raptorx.uchicago.edu/>) (Xu et al., 2021). 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 (Wiederstein & Sippl, 2007). 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 (Wiederstein & Sippl, 2007). 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/>) (Zhou et al., 2011).

### Assessment of Model Quality

Subsequently, the predicted three-dimensional structure was evaluated using PROCHECK (<https://servicesn.mbi.ucla.edu/PROCHECK/>), Verify3D (<http://nihserver.mbi.ucla.edu/Verify3D/>), and ERRAT Structure Evaluation server (<https://servicesn.mbi.ucla.edu/ERRAT/>).

### 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 (Szklarczyk et al., 2015).

### Disulfide Bonding in the Protein

The formation of disulfide bonds between cysteine residues in a protein is critical for its folding into a functional and stable shape. To get insight into the experimental structure determination and stability of a hypothetical protein, we employed CysPred and DIANA to predict disulfide bonds within the protein. CysPred (<https://gpcr.biocomp.unibo.it/cgi/predictors/cyspred/predcyspred.cgi>) 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 (Grützner et al., 2009). DIANA

(<https://clavius.bc.edu/clotelab/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 (Ferrè & Clote, 2005). 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

Galaxy server (<http://galaxy.seoklab.org/index.html>) was used for ligand-protein binding site prediction in hypothetical protein. 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 (Heo et al., 2014).

### Detecting Active Sites

The Computed Atlas of Surface Topography of Protein (CASTp) (<http://sts.bioengr.uic.edu/castp/>) was used to find the active site of the protein (Dundas et al., 2006). It is a web-based tool for identifying, defining, and quantifying concave surface areas on 3D protein structures.

### Molecular Docking and Simulation Studies

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 protein-protein interactions (Kozakov et al., 2017). 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 (Laskowski et al., 2018).

HawkDock was used to calculate the binding energy of a protein-protein complex (Weng et al., 2019). To estimate interaction free energies between diverse protein-protein complexes, the service uses molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) (Weng et al., 2019).

## Result and Discussions

### Similarity Identification, and Phylogeny Analysis

The results of BLASTp against a non-redundant database revealed similarities with other flagellum proteins (Table 1). The FASTA sequences of the hypothetical protein (HP) (WP\_058430053.1) and homologous identified proteins were aligned using multiple sequence alignment (Figure 1). Phylogenetic analysis was used to corroborate the homology assessment between the proteins, down to the complex and subunit level. The alignment and BLAST results were used to create a phylogenetic tree, which offers a comparable idea about the protein (Figure 1). The distances between branches are also taken into consideration.

### Physicochemical Features

The protein consists of 384 amino acids, among the most abundant was Leu (L) 50 followed by, Ala (A) 35, Asn (N) 28, Lys (K) 27, Glu (E) 25, Thr (T) 25, Ser (S) 24,

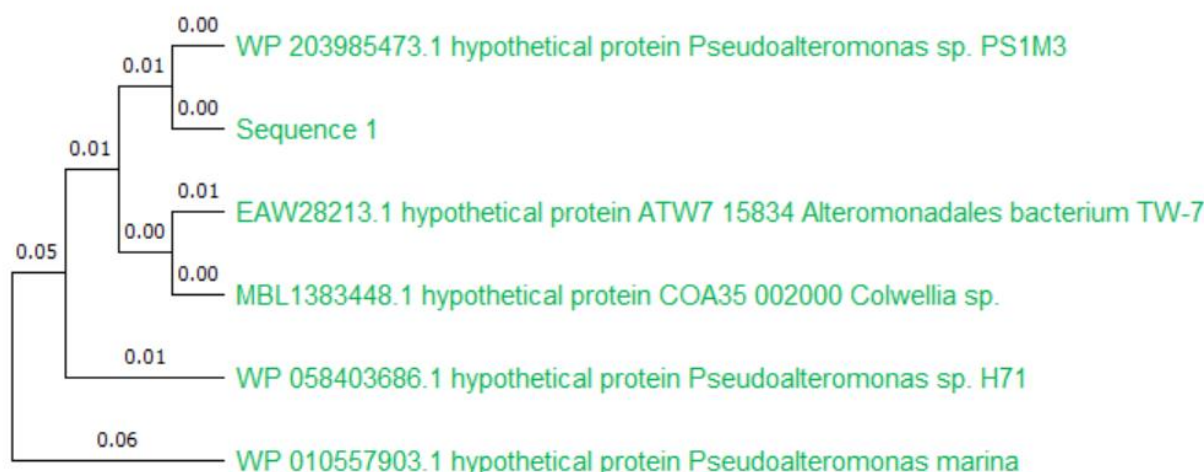
Thr (T) 25, Glu (E) 25, Tyr (Y) 24, Val (V) 22, Gln (Q) 21, Phe (F) 17, Arg (R) 16, Asp (D) 15, Ile (I) 15, Pro (P) 12, His (H) 7, Met (M) 5, and Cys (C) 2. The computed molecular weight was 43663.71 Da, with a theoretical pI of 8.50, indicating a positively charged protein. The total number of positively charged (Arg + Lys) and negatively charged (Asp + Glu) residues were discovered to be 43 and 40, respectively. The protein was classified as stable by the computed instability index of 31.22 according to the previous study (Islam et al., 2022). The aliphatic index was 91.74, indicating that proteins are stable across a wide temperature range (Jahan et al., 2021). The GRAVY value was -0.336. 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<sub>1967</sub>H<sub>3086</sub>N<sub>522</sub>O<sub>587</sub>S<sub>7</sub>.

### Functional Annotation of the Hypothetical Protein

This potential protein sequence was discovered to have only a domain using the conserved domain search tool which is PEP\_TPR\_lipo superfamily (accession No. cl37187). This protein family occurs strictly within a subset of Gram-negative bacterial species with the proposed PEP-CTERM/exosortase system, analogous to the LPXTG/sortase system common in Gram-positive bacteria (Haft et al., 2006). This protein occurs in a species if and only if a transmembrane histidine kinase

**Table 1.** Similar protein obtained from non-redundant sequences

Protein ID	Accession No. WP_058430053.1			
	Organism	Protein Name	Identity (%)	e value
WP_203985473.1	<i>Pseudoalteromonas sp. PS1M3</i>	hypothetical protein	100	0.00
EAW28213.1	<i>Alteromonadales bacterium TW-7</i>	hypothetical protein ATW7_15834	98.43	0.00
MBL1383448.1	<i>Colwellia sp.</i>	sodium:solute symporter	98.17	0.00
WP_058403686.1	<i>Pseudoalteromonas sp. H71</i>	hypothetical protein	97.14	0.00
WP_010557903.1	<i>Pseudoalteromonas marina</i>	hypothetical protein	89.32	0.00



**Figure 1.** Multiple sequence phylogenetic tree.

(TIGR02916) and a DNA-binding response regulator (TIGR02915) also occur. The presence of tetratricopeptide repeats (TPR) suggests protein-protein interaction, possibly for the regulation of PEP-CTERM protein expression, since many PEP-CTERM proteins in these genomes are preceded by a proposed DNA binding site for the response regulator (Haft et al., 2006). Two further domain search tools, InterProScan and Pfam, were used to verify the result and both predicted no domain.

### 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 OuterMembrane (Table 3). The number of transmembrane helices predicted by TMHMM and HMMTOP was zero. All of these findings point to the protein being OuterMembrane. Furthermore, the subcellular localization of this HP will help to design drugs or vaccines against this bacterial pathogen (Islam et al., 2022; Jahan et al., 2021).

### Secondary Structure Analysis

The proportions of alpha helix, beta-sheet content, coil content, and overall confidence value were 74 %, 0 %, 26 %, and 81.5%, respectively, according to the PROTEUS Structure Prediction Server 2.0 study. Secondary structures have a crucial role in the structure

and folding of proteins. It is fascinating to study proteins in terms of their highly regular secondary structure motifs (Islam & Jahan, 2022; Ji & Li, 2010).

### Three-dimensional Structure Prediction, Model Quality Refinement, and Assessment

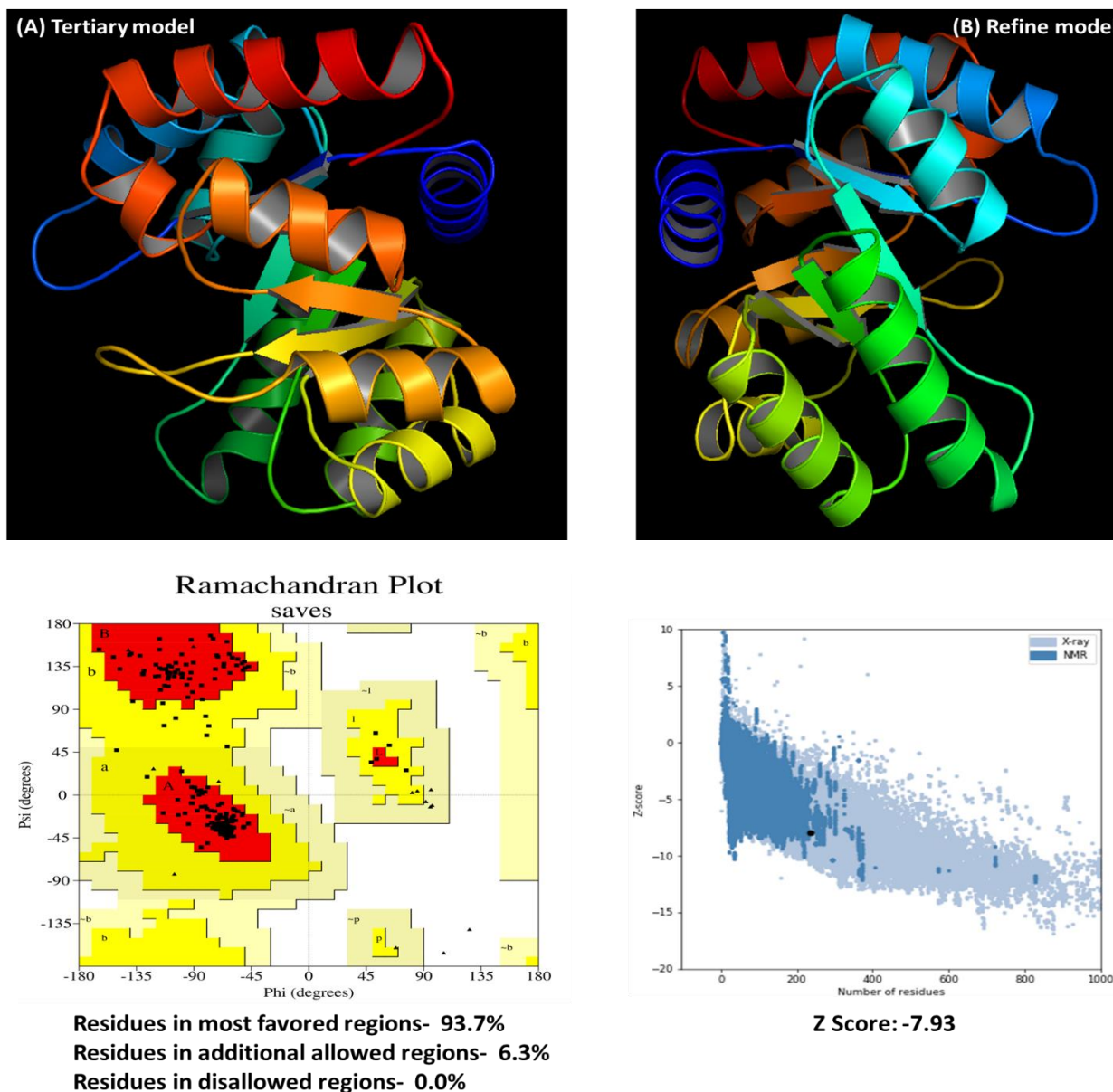
The three-dimensional structure of the target protein was predicted using the RaptorX server (<http://raptorx.uchicago.edu/>) and protein model 1 was chosen. The RaptorX program predicts 3D structures for protein sequences that have no close homologs in the Protein Data Bank (PDB) developed by the Xu group. A sequence input is used to predict secondary and tertiary structures, solvent accessibility, disordered regions, and solvent accessibility, according to RaptorX (Källberg et al., 2014). 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). Validation of the galaxy server refined model was assessed by PROCHECK through Ramachandran plot analysis, where the distribution of  $\phi$  and  $\psi$  angle in the model within the limits are shown (Figure 2C). A valid model covers 93.7% 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

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

ProtParam tool		EMBOSS Pepstats	
Sequence ID	WP_058430053.1	Charge	6.5
Family (Pfam)	undefined	Improbability of expression in inclusion bodies	0.682
Domain (ScanProsite)	5 hits	Average residue weight	113.708
Alignment	21– 403	A280 extinction coefficients 1 mg ml <sup>-1</sup>	0.819
HMM length	383	A280 molar extinction coefficients	35760
Bit score	9.446	Tiny (A + C + G + S + T)	100
E value	2.37 × 10 <sup>-102</sup>	Small (A + B + C + D + G + N + P + S + T + V)	177
Number of AA	384	Aliphatic (A + I + L + V)	122
MW	43663.71 Da	Aromatic (F + H + W + Y)	48
pI	8.50	Non-polar (A + C + F + G + I + L + M + P + V + W + Y)	196
Extinction coefficients	35885	Polar (D + E + H + K + N + Q + R + S + T + Z)	188
Instability index	31.22	Charged (B + D + E + H + K + R + Z)	90
Aliphatic index	91.74	Basic (H + K + R)	50
GRAVY	-0.336	Acidic (B + D + E + Z)	40

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

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



**Figure 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.

model. On the Verify3D graph, 91.57% of residues have an average 3D-1D score of  $\geq 0.2$ , showing that the model has an excellent environmental profile, and an overall quality factor of 92.9515 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  $-339,331.5$  kJ/mol, making the modeled structure more stable. In addition, ProSA web server analysis resulted in a Z score of  $-7.93$  which indicates the model validation (Islam et al., 2022)(Figure 2D).

**Protein-protein Interaction Analysis**

The STRING 10.0 search was used to find a probable functional interaction network for the protein

[31]. The identified functional partners with scores were; Alanyl-tRNA ligase (0.897), Hypothetical protein (0.895), COG0639 Diadenosine tetraphosphatase and related serine/threonine protein phosphatases (0.885), Molecular chaperone groel (0.871), Chaperone protein hsca (0.785), Chaperone (0.785), Molecular chaperone dnak (0.785), Heat shock protein 90(HSP90) (0.776) and Hypothetical protein (0.764) (Figure 3). Almost all cellular processes involve protein-protein interactions (PPIs). These interactions play an important role in controlling signaling pathways (Moreira et al., 2006).

**Ligand Binding Interactions**

The PDB file of the best-predicted domain-A model was submitted to the Galaxy server for ligand binding site predictions by matching target models. Three

models were predicted by a galaxy server with different ligands. Galaxy server also combines the results into three parts such as Predicted ligand-binding residues, Predicted binding poses of the model, and Templates for protein-ligand complex (Table 5; Figure 4 (A)). 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 were given in Figure 4. Cooperation between competing biological functions is essential for the regulation of ligand binding. Molecular mechanisms involve protein conformational transitions between low- and high-affinity states to regulate cellular processes via cooperative protein-ligand interactions (Moreira et al., 2006).

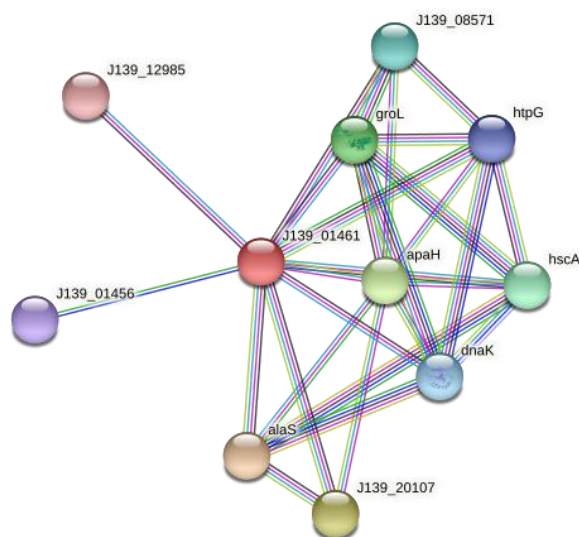
**Active Site Detection**

The CASTp v.3.0 algorithm predicted 5 unique active sites in the modeled protein (Figure 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. The top active sites of the modeled protein were identified between the area of 67.142 and the volume of 51.783 (Figure 5). Figure 5

shows the protein's anticipated active site together with its amino acid residues. An active site contains a binding site that binds the substrate and orients it for catalysis. In some cases, the substrate's orientation and its proximity to the active site contribute so much to the enzyme's ability to function, that even when all other parts have been mutated and lost their functionality, the enzyme can still perform. So it is indeed very important to find the active site of a protein (Islam & Jahan, 2022; Jahan et al., 2021).

**Molecular Docking and Simulation Studies Between Hypothetical Protein and Heat Shock Protein 90 (HSP90)**

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 HSP90 and amino acid sequence of the hypothetical protein. The amino acid interactions revealed that the complex is stabilized by 4 hydrogen bonds (Figure 7B). Within the HSP90-hypothetical protein complex, Gln199, Tyr232, Tyr232, and Pro163 of HSP90 protein respectively interact with Asn346, Arg374, Asn348, and Ser319 of HSP 90 protein through H-bonds (Figure 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 – 50.72 (kcal/mol) proving that the complex is stable.

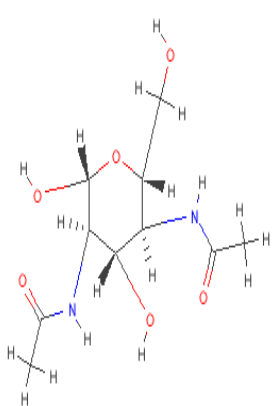
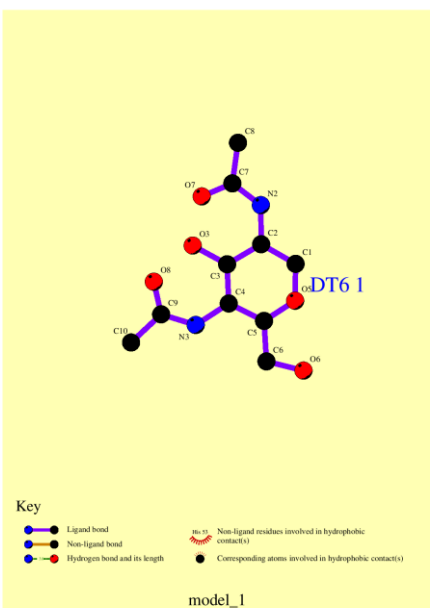


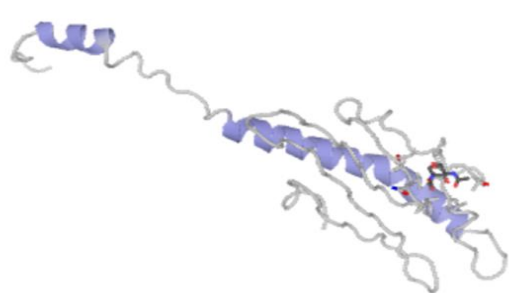
**Figure 3.** String (Protein-protein interactions) analysis of hypothetical protein.

**Table 4.** Prediction of cysteine residues involved in disulphide bonding by CYPRED and DIANA

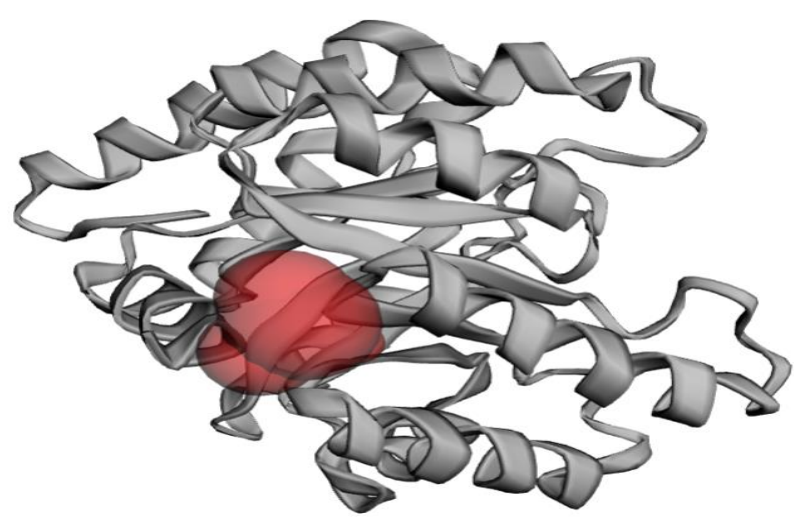
Cysteine	CYPRED			DIANA
	Prediction	Reliability	Distance	Bonded cysteine
CYS 17	NON-Bonding State	7		
CYS 111	NON-Bonding State	3	11	QEWKECPDYVS -AGENSCYFNSS

**Table 5.** Predicted ligand-binding residues

No	Ligand Name	Ligand Structure	Binding residues	Interaction analysis
1	DT6		71L 72S 73G 79P 80D 93V 94N 95A 96A	



**Figure 4.** Predicted binding poses.



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



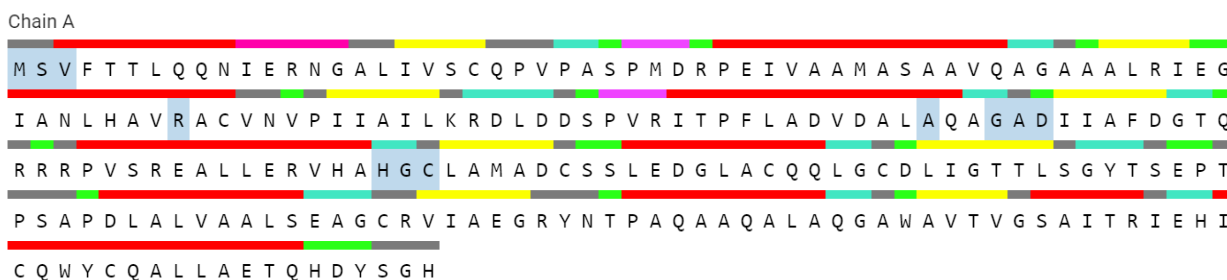


Figure 6: Active location of the hypothetical protein. The active site of amino acid residues (Blue color).

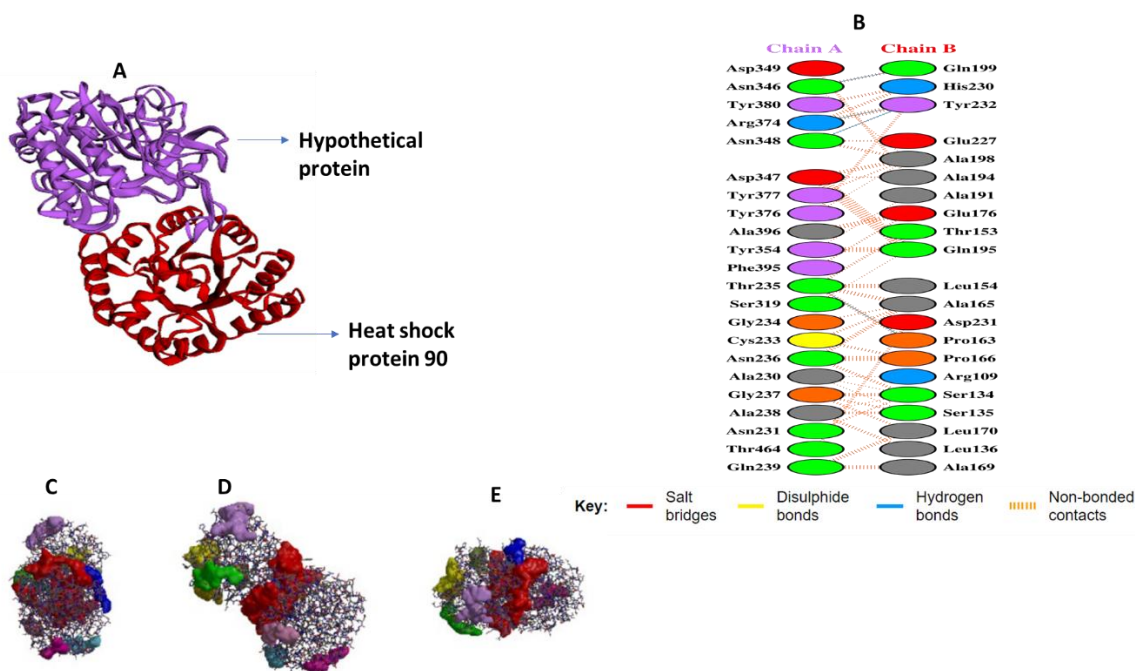


Figure 7: (A) HSP90-hypothetical protein complex, (B) Interactions between the protein-protein complex and (C, D, E) Three different binding sites of the selected protein-protein complex.

Additionally, three different binding sites of the protein complex was shown in Figure 7 (C, D, E). HSP90 can be an interesting tool for studies of adaptation and evolution because it interacts with essential and hypothetical proteins, is ubiquitous, and is highly conserved. It is a conserved molecular chaperone that works as part of complexes in which it is targeted to a variety of substrates by distinct client proteins (García-Descalzo et al., 2011). Various bacterial organisms from extreme environments have been thoroughly characterized and increased access to genome sequence and characterization of many protein data has lately enabled the identification of stress-related protein families (Laksanalamai & Robb, 2004). Moreover, through this work, we found that hypothetical protein from *Pseudoalteromonas spp.* interact strongly with HSP90 which indicates the common adaptation pattern of bacterial species to the marine environment, leading to positive covariation and explaining the long-term stability of seaweed population and abundance in the absence of compensatory dynamics.

### 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 that can be a biomarker. The existence and distribution of this hypothetical protein domain across a wide range of bacterial strains and interaction with HSP90 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 HSP90 protein are found in a stable interaction therefore it may be of interest to researchers looking to produce new drugs against *Pseudoalteromonas spp* for the treatment of diseases in seaweed.

### Ethical Statement

Not applicable.

## Funding Information

There is no funding institution for this study.

## Author Contribution

Conceptualization, **M.M.** and **SK.**; methodology, **SK.** and **M.M.** and **S.S.**; software, **SK.**; validation, **M.M.**, and **SK.**; formal analysis, **SK.**; investigation, **SK.**; resources, **SK.**; data curation, **M.M.** and **SK.**; writing—original draft preparation, **M.M.**; writing—review and editing, **SK.** and **S.M.** All authors have read and agreed to the published version of the manuscript.

## Conflict of Interest

The author(s) declare that they have no known competing financial or non-financial, professional, or personal conflicts that could have appeared to influence the work reported in this paper.

## Acknowledgements

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.

## References

- Alzohairy, A. (2011). BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences*, 2, 60-61.
- Bhasin, M., Garg, A., & Raghava, G. P. S. (2005). PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics*, 21(10), 2522-2524. <https://doi.org/10.1093/bioinformatics/bti309>
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., & Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, 31(1), 365-370. <https://doi.org/10.1093/nar/gkg095>
- Dobson, L., Reményi, I., & Tusnády, G. E. (2015). CCTOP: a Consensus Constrained TOPology prediction web server. *Nucleic Acids Research*, 43(W1), W408-W412. <https://doi.org/10.1093/nar/gkv451>
- 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. <https://doi.org/10.1093/nar/gkl282>
- Ferrè, F., & Clote, P. (2005). DIANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res*, 33(Web Server issue), W230-232. <https://doi.org/10.1093/nar/gki412>
- Finn, R. D. (2005). Pfam: the protein families database. In *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*. <https://doi.org/https://doi.org/10.1002/047001153X.g306303>
- García-Descalzo, L., Alcazar, A., Baquero, F., & Cid, C. (2011). Identification of in vivo HSP90-interacting proteins reveals modularity of HSP90 complexes is dependent on the environment in psychrophilic bacteria. *Cell Stress Chaperones*, 16(2), 203-218. <https://doi.org/10.1007/s12192-010-0233-7>
- 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. <https://doi.org/10.1093/nar/gkg563>
- 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. <https://doi.org/10.1016/j.bpj.2009.05.037>
- Haft, D. H., Paulsen, I. T., Ward, N., & Selengut, J. D. (2006). Exopolysaccharide-associated protein sorting in environmental organisms: the PEP-CTERM/EpsH system. Application of a novel phylogenetic profiling heuristic. *BMC Biology*, 4(1), 29. <https://doi.org/10.1186/1741-7007-4-29>
- 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. <https://doi.org/10.1093/nar/gku321>
- Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Mulder, N., Natale, D., Orengo, C., Quinn, A. F., Selengut, J. D., Sigrist, C. J. A., Thimma, M., Thomas, P. D., Valentin, F., Wilson, D., Wu, C. H., & Yeats, C. (2008). InterPro: the integrative protein signature database. *Nucleic Acids Research*, 37(suppl\_1), D211-D215. <https://doi.org/10.1093/nar/gkn785>
- Islam, S., & Jahan, M. (2022). Functional Annotation of Uncharacterized Protein from Photobacterium damsela subsp. piscicida (Pasteurella piscicida) and Comparison of Drug Target Between Conventional Medicine and Phytochemical Compound Against Disease Treatment in Fish: An In-silico Approach. *Genetics of Aquatic Organisms*, 6, 453. <https://doi.org/10.4194/GA453>
- Islam, S., Sanjida, S., Jahan, M., Mahfuj, M. s. E., & 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. <https://doi.org/10.38001/ijlsb.1032171>
- Jahan, M., Islam, S., & Mahfuj, M. s. E. (2021). In Silico Functional Annotation of VP 128 Hypothetical Protein from Vibrio parahaemolyticus. <https://doi.org/10.4194/AFS37>
- Ji, Y.-Y., & Li, Y.-Q. (2010). The role of secondary structure in protein structure selection. *The European Physical Journal E*, 32(1), 103-107. <https://doi.org/10.1140/epje/i2010-10591-5>
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, 36(suppl\_2), W5-W9. <https://doi.org/10.1093/nar/gkn201>
- Källberg, M., Margaryan, G., Wang, S., Ma, J., & Xu, J. (2014). RaptorX server: a resource for template-based protein structure modeling. *Methods in molecular biology (Clifton, N.J.)*, 1137, 17-27. <https://doi.org/10.1007/978->

- 1-4939-0366-5\_2
- Kanehisa, M., Goto, S., Kawashima, S., & Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Research*, 30(1), 42-46. <https://doi.org/10.1093/nar/30.1.42>
- Kim, J., Yarish, C., Hwang, E. K., Park, M., & Kim, Y. (2017). Seaweed aquaculture: Cultivation technologies, challenges and its ecosystem services. *ALGAE*, 32, 1-13. <https://doi.org/10.4490/algae.2017.32.3.3>
- 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. <https://doi.org/10.1038/nprot.2016.169>
- Laksanalamai, P., & Robb, F. T. (2004). Small heat shock proteins from extremophiles: a review. *Extremophiles*, 8(1), 1-11. <https://doi.org/10.1007/s00792-003-0362-3>
- Largo, D. B., Fukami, K., & Nishijima, T. (1995). Occasional pathogenic bacteria promoting ice-ice disease in the carrageenan-producing red algae *Kappaphycus alvarezii* and *Euclima denticulatum* (Solieriaceae, Gigartinales, Rhodophyta). *Journal of Applied Phycology*, 7(6), 545-554. <https://doi.org/10.1007/BF00003941>
- 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. <https://doi.org/10.1002/pro.3289>
- Li, J., Majzoub, M. E., Marzinelli, E. M., Dai, Z., Thomas, T., & Egan, S. (2022). Bacterial controlled mitigation of dysbiosis in a seaweed disease. *The ISME Journal*, 16(2), 378-387. <https://doi.org/10.1038/s41396-021-01070-1>
- Lim, P. E., Yang, L.-E., Tan, J., Maggs, C., & Brodie, J. (2017). Advancing the taxonomy of economically important red seaweeds (Rhodophyta). *European Journal of Phycology*, 52, 438-451. <https://doi.org/10.1080/09670262.2017.1365174>
- 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., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Marchler, G. H., Mullokandov, M., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Yamashita, R. A., Yin, J. J., Zhang, D., & Bryant, S. H. (2005). CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Research*, 33(suppl\_1), D192-D196. <https://doi.org/10.1093/nar/gki069>
- McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4), 404-405. <https://doi.org/10.1093/bioinformatics/16.4.404>
- Möller, S., Croning, M. D. R., & Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, 17(7), 646-653. <https://doi.org/10.1093/bioinformatics/17.7.646>
- Moreira, I. S., Fernandes, P. A., & Ramos, M. J. (2006). Unraveling the importance of protein-protein interaction: application of a computational alanine-scanning mutagenesis to the study of the IgG1 streptococcal protein G (C2 fragment) complex. *J Phys Chem B*, 110(22), 10962-10969. <https://doi.org/10.1021/jp054760d>
- 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. <https://doi.org/https://doi.org/10.1016/j.jtbi.2008.10.07>
- Shinn, A., Pratoomyot, J., Bron, J., Paladini, G., Brooker, E., & Brooker, A. (2015). Economic costs of protistan and metazoan parasites to global mariculture. *Parasitology*, 142, 196-270. <https://doi.org/10.1017/S0031182014001437>
- Syafitri, E., Prayitno, S., Ma'ruf, W., & Radjasa, o. k. (2017). Genetic diversity of the causative agent of ice-ice disease of the seaweed *Kappaphycus alvarezii* from Karimunjawa island, Indonesia. *IOP Conference Series: Earth and Environmental Science*, 55, 012044. <https://doi.org/10.1088/1755-1315/55/1/012044>
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., & von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*, 43(Database issue), D447-452. <https://doi.org/10.1093/nar/gku1003>
- Tsiresy, G., Preux, J., Lavitra, T., Dubois, P., Lepoint, G., & Eeckhaut, I. (2016). Phenology of farmed seaweed *Kappaphycus alvarezii* infestation by the parasitic epiphyte *Polysiphonia* sp. in Madagascar. *Journal of Applied Phycology*, 28. <https://doi.org/10.1007/s10811-016-0813-8>
- Tusnády, G. E., & Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics*, 17(9), 849-850. <https://doi.org/10.1093/bioinformatics/17.9.849>
- Ward, G., Faisan, J., Cottier, E., Gachon, C., Hurtado, A., Lim, P., Matoju, I., Msuya, F., Bass, D., & Brodie, J. (2019). A review of reported seaweed diseases and pests in aquaculture in Asia. *Journal of the World Aquaculture Society*, 51. <https://doi.org/10.1111/jwas.12649>
- 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. <https://doi.org/10.1093/nar/gkz397>
- 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. <https://doi.org/10.1093/nar/gkl910>
- 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. <https://doi.org/10.1038/s42256-021-00348-5>
- Yu, C., & Hwang, J. (2008, 26-28 Nov. 2008). Prediction of Protein Subcellular Localizations. 2008 Eighth International Conference on Intelligent Systems Design and Applications,
- 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., & Brinkman, F. S. L. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, 26(13), 1608-1615. <https://doi.org/10.1093/bioinformatics/btq249>