

# 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

Sk Injamamul Islam<sup>1,\*</sup> , Mou Moslema Jahan<sup>2</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.

## How to cite

Islam, S.I., Jahan, M.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(3.Special Issue), GA453. <http://doi.org/10.4194/GA453>

## Article History

Received 24 August 2021  
Accepted 17 January 2022  
First Online 18 January 2022

## Corresponding Author

Tel.: +0646904801  
E-mail:  
injamamulislam017@gmail.com

## Keywords

Phytochemical screening  
Drug target  
ADMET  
Molecular docking  
*Photobacterium damsela*

## Abstract

*Photobacterium damsela* is one of the deadliest bacterial pathogens in the world of aquaculture. *P. damsela* has an uncharacterized protein (UniprotKb ID: A0PB24) that can enter the host cell and induce infection in fish, according to research. Domains of unknown function are proteins that have been identified experimentally but do not have a known functional or structural domain. Targeting this uncharacterized protein, we found that this protein may have link with other virulent proteins of *P. damsela*. So, the main objective of the study was to reveal the functional annotation of the protein and compare the inhibitory potentiality of conventional drug (Flumequine, Oxytetracycline and Florfenicol) and four potential antibacterial compounds from *Syzygium aromaticum* which reported previously. The uncharacterized protein was modeled, refined, and validated in this experiment. The protein-protein interaction network suggested that this uncharacterized protein may serve as a DNA/RNA hydrazase and conserved domain search predict two different domains. Furthermore, the binding affinity of conventional drugs and phytochemical compounds were calculated against the protein with the docking approaches and predict *S. aromaticum* inhibitory compounds bind strongly with the protein. In addition, the ADMET properties of the compounds confirmed the drug likeness properties of the four compounds.

## Introduction

Photobacteriosis (formerly *Pasteurella piscicida*) is a bacterial septicaemia caused by the halophilic bacterium *Photobacterium damsela* subsp. *piscicida* belong to the vibronaceae family (Shieh et al., 2003). This bacterium has been identified as a pathogen in a range of aquatic life, including crustaceans, mollusks, fish, and crustaceans and result in harmful bacterial diseases impacting both wild and cultured marine fish species in Europe, Asia, and the Mediterranean region (Toranzo & Barja, 1993; Andreoni & Magnani, 2014).

Furthermore, this bacterial pathogen has been documented to cause infections in people, suggesting that it could be a zoonoses agent. Several studies have been conducted over the last 20 years to determine the effectiveness of immunization in preventing pasteurellosis and controlling the disease. Previously, antibacterial drugs such as oxytetracycline, flumequine, and florfenicol were used to combat the disease outbreak, but no drug has shown consistent success (Zhang et al., 2020). Formulations based on LPS and ribosome fractions of bacteria produced better outcomes. However, these formulations faced not just

repeatability issues, but also production challenges on a wide scale. Passive vaccination was also tested, but because to the short duration of protection, it was only successful if the vaccine was given up to 24 hours before the challenge.

Other fish infections, such as *Vibrio anguillarum*, *Aeromonas salmonicida*, and the clinically significant *Pasteurella* spp., *Pasteurella multocida* and *Pasteurella haemolytica*, have been identified to share major phenotypic traits with *P. piscicida* (Magariños et al., 1992). The existence of granulomas in various internal organs such as the liver, kidney, and spleen, which vary in size depending on their age, are pathognomonic lesions linked with this disease in chronic cases (Mladineo et al., 2006). The tremendous expansion of ISs resulted in the emergence of a large number of pseudogenes, which resulted in a significant loss of biological functions. Pseudogene development is one of the key aspects of the *Pdp* genome, and it accounts for the majority of the ecological and phenotypic differences between it and its sibling subspecies, *P. damsela subsp. damsela*, as well as other Vibrionaceae (Balado et al., 2017).

Uncharacterized protein families (UPFs) are one type of hypothetical protein. Unknown proteins have experimental structures that have been discovered but not defined or connected to a known gene. Analyzing the function of proteins with unknown functions has a number of advantages, including the ability to determine new conformational orientations of 3-dimensional structures, which allows for the evaluation of new domains and motifs, as well as the discovery of new protein mechanisms and cascades (Lubec et al., 2005). Uncharacterized proteins based on nucleic acid sequences that require experimental protein chemistry data. Furthermore, these proteins have a low level of similarity to known, annotated proteins (Lubec et al., 2005).

Medicinal plants can play a critical role in the treatment of a variety of ailments, particularly in areas where resources are scarce., traditional drugs based on plants have less detrimental consequences than modern drugs, which is one of the main reasons why essential chemicals are extracted and produced from plants (Abushouk et al., 2017). *Syzygium aromaticum* is a typical plant whose importance has risen steadily in recent years around the world. It contains a large number of biologically active compounds with a variety of structures. *S. aromaticum* (clove) is a traditional spice that has been used for food preservation and possesses various pharmacological activities. *S. aromaticum* is rich in many phytochemicals as follows: Syzyginin A, Syzyginin B, Casuarictin, hydrocarbon, and phenolic compounds. *S. aromaticum* has been tested pharmacologically against a variety of pathogenic bacteria, parasites, and microbes, including pathogenic bacteria, Plasmodium, Babesia, Herpes simplex, and hepatitis C viruses. Clove oil has been shown to have analgesic, antioxidant, anticancer, antispasmodic,

antiviral, antifungal, and antibacterial activity against a variety of pathogenic bacteria, including methicillin-resistant *Staphylococcus epidermidis* and *S. aureus*, according to several studies (Batiha et al., 2020).

To introduce effective medicines in a conventional or standard manner can take a long time, be expensive, and require a significant amount of effort (Lim et al., 2015). For example, high-throughput screening (HTS) is a technique that integrates multiple-well microplate with automated processing to improve drug development by assaying a large number of putative drug-like molecules (Hughes et al., 2011). Additionally, HTS should have abundant resources, as processing a particular HTS program is expensive and involves the use of robotic devices (Szymański et al., 2012). On the contrary, computer-aided drug design, also known as in silico drug design, is a relatively new technology for screening a large database of compounds using a high-throughput approach (Liang et al., 2006). The in silico virtual screening approach aids in the discovery of novel medicines by generating hits for lead compounds in a shorter period and at a cheaper cost (Wichapong et al., 2013). As a result, improved in silico drug design reduces the time required to develop, design, and optimize a novel drug. The virtual screening approach has been used for decades to find the best lead compounds with various structural properties for use with a given biological target (J. Wu et al., 2020). Furthermore, computer-aided drug design has been used to find a wide variety of interesting drug applications and hits utilizing virtual screening and molecular docking (C. Wu et al., 2020). In light of the above-mentioned *S. aromaticum* drugs, the goal of this study is to use molecular docking to screen active compounds of *S. aromaticum* against the uncharacterized protein (UniProtKb: AOPB24) and investigate their interaction pattern. As a result, virtual screening, molecular docking, and ADMET (absorption, distribution, metabolism, excretion, and toxicity) features strategies were combined to screen potential natural anti-fish drugs in this study.

## Materials and Methods

### Selection and Retrieving the Sequence of Uncharacterized Protein

The UniProtKb database (<https://www.uniprot.org/>) was used to retrieve the amino acid (aa) sequence of the uncharacterized protein (UniProtKb: AOPB24) found in *Photobacterium damsela subsp. piscicida* (*Pasteurella piscicida*) and downloaded in FASTA format.

### Similarity Identification

By using the BLASTp program, the NCBI protein database (<http://www.ncbi.nlm.nih.gov/>) was searched to find proteins that had structural similarities to the

uncharacterized protein to get preliminary information about its function (Madden et al., 2019).

### Multiple Sequence Alignment and Phylogeny Analysis

Multiple sequence alignments between the uncharacterized protein and proteins that had structural similarities with the uncharacterized protein were performed using the BioEdit biological sequence alignment editor program (Alzohairy, 2011). The phylogenetic analysis was performed using Molecular Evolutionary Genetic Study (MEGA) version 10.0.5 (<https://megasoftware.net/>).

### Conserved Domain, Motif and Fold Identification

Domain search carried out at conserved domain database (CDD, available at NCBI) (Marchler-Bauer & Bryant, 2004), for conserved domain. The Motif (Genome Net) server was used to find protein motifs. Pfam were used to determine the evolutionary connections of the protein. Protein sequence analysis and classification server InterProScan was employed for the functional analysis of the protein. For protein folding pattern recognition, PFP-FunD SeqE server was used.

### Physicochemical Characterization of the Protein

The physicochemical properties of the hypothetical protein in raw sequence format were assessed using the ProtParam tool from the ExPASy service (Wilkins et al., 1999). The tool computes and provides the molecular weight, theoretical pI, amino acid composition, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index, and grand average of hydropathicity (GRAVY), among other metrics. The extinction coefficient of a protein is a measure of how much light it absorbs at a given wavelength. The instability index is a measurement of a protein's stability in a test tube. An instability index of less than 40 indicates that the situation is stable, whereas a value more than 40 indicates that the situation is unstable. The relative volume filled by aliphatic side chain amino acids is defined as a protein's aliphatic index. The GRAVY value of a peptide or protein is computed by multiplying the total of all amino acid hydrophobicity values by the number of residues in the sequence (<http://web.expasy.org/protparam/protparam-doc.html>).

### Assessment of Secondary Structure

The secondary structural elements of the protein (UniprotKb: A0PB24) were predicted through the SOPMA tool (Combet et al., 2000) using the default parameters (window width of 17, number of states of 4, and similarity threshold of 8).

### Sub-cellular Localization

It is necessary to obtain information on a protein's subcellular localization in order to predict its cellular function. The outer membrane, inner membrane, periplasm, extracellular space, and cytoplasm all include proteins (Gazi et al., 2016). CELLO was used to predict bacterial protein subcellular localization (Shen & Chou, 2007).

### HHPred Model Generation

HHpred is a fast server that detects protein homology and predicts protein structure. It is the first server that compares profile hidden Markov models (HMMs) in a pairwise manner (Söding et al., 2006). It searches a wide variety of databases, including PDB, SCOP, Pfam, SMART, COGs, and CDD whereas conventional sequence search methods examine sequence databases, such as UniProt non-redundant databases. HHpred is also a fast server that uses hidden Markov model pairwise comparison profiles (HMMs) to discover and forecast remote protein homology and structure.

### Refinement, and Validation of Three-Dimensional Structures

GalaxyWeb refined the 3D structure of the protein. The validity of the structure is a critical step in homology modeling, which is based on empirically validated 3D protein structures. For basic confirmation, the proposed protein model was submitted to ProSA-web (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/>).

### SOSUI Server

SOSUI is a free online tool that determines whether an amino acid sequence supplied is a membrane protein and distinguishes between insoluble and soluble proteins. The algorithm considers four physicochemical parameters: The Kyte and Doolittle hydrophobicity index, an amphiphilicity index, an amino acid charge index, and the length of each sequence. The kind of protein, prediction of transmembrane helices for membrane proteins, graph of the hydrophobicity plot, and helical wheel diagrams of all transmembrane helices are among the output outputs (Hirokawa et al., 1998).

## Model Quality Assessment

PROCHECK was used to evaluate the quality of the anticipated three-dimensional structure (Laskowski et al., 1993), Verify3D ([http://nihserver.mbi.ucla.edu/Verify\\_3D/](http://nihserver.mbi.ucla.edu/Verify_3D/)) (Eisenberg et al., 1997) and ERRAT Structure Evaluation server (Colovos & Yeates, 1993).

## Protein-Protein Interaction Analysis

Protein residues interact with one another in order to perform their functions correctly. STRING (<http://string-db.org/>), in this study, we used a database of known and anticipated protein interactions that uses physical and functional connections. This was based on Genomic Context, high-throughput studies, (Conserved) Co-expression, and Prior Knowledge. This database integrates interaction data from the following sources quantitatively (Franceschini et al., 2013).

## Active Site Detection

The Computed Atlas of Surface Topography (CASTp) of Proteins was used to determine the active site of the protein (<http://sts.bioengr.uic.edu/castp/>) (Dundas et al., 2006) provides an online resource for locating, identifying and quantifying concave surface areas on three-dimensional protein structures.

## Preparation of Protein

The protein's 3D structure was modelled and developed using the following criteria: water, metal ions, and cofactors were removed, polar hydrogen atoms were introduced, nonpolar hydrogen was combined, and gasteiger charges were calculated using AutoDockTools.

## Retrieval and Preparation of Compounds

Phytochemicals derived from naturally occurring medicinal plants cover a wide range of chemical spaces that can be used in drug development and discovery. IMPPAT stands for Indian Medicinal Plants, Phytochemistry, and Therapeutics, a manually curated database of over 1742 Indian medicinal plants and over 9500 phytochemical compounds that uses cheminformatic methodologies to improve natural product-based drug discovery (Mohanraj et al., 2018). Because of virtual screening, the phytochemical of the *Syzygium aromaticum* has been discovered and obtained from the database. The compounds found from the database were created by assigning accurate AutoDock 4 atom types, merging nonpolar hydrogens, detecting aromatic carbons, and establishing a 'torsion tree. Researchers have found that the majority of atoms in the compound have the same type of atom as the AD4 atom (Seeliger & de Groot, 2010).

## Molecular Docking and Receptor Grid Generation

The PyRx virtual screening tool AutoDock Vina was used to create a protein receptor grid (Dallakyan & Olson, 2015). The molecular docking investigation was carried out using the PyRx virtual screening program AutoDock Vina to find the binding mechanism of the required protein with chosen phytochemicals. PyRx is an open-source virtual screening application that can screen libraries of compounds against a given therapeutic target and is primarily used in CADD techniques. PyRx integrates AutoDock 4 and AutoDock Vina as docking wizards with an intuitive user interface, making it a more trustworthy CADD tool. This experiment used PyRx's AutoDock Vina wizard for molecular docking to find the optimum protein and ligand binding poses. For docking objectives, the default configuration parameters of the PyRx virtual screening tools were utilized, and the highest binding energy (kcal/mol) with the negative sign was chosen for further investigation. Subsequently, using the BIOVIA Discovery Studio Visualizer v19.1.0.18287, the binding interaction of the protein–ligands complex was seen.

## Absorption, Distribution, Metabolism, and Excretion (ADME) and Toxicity Test

The physicochemical, pharmacokinetics, metabolism, and excretion properties of molecules into urine and feces are all listed in the ADME of a substance (Y. Li et al., 2019). The Swiss-ADME server (<http://www.swissadme.ch/>) was used to forecast the various pharmacokinetic and pharmacodynamic parameters for the experiments (Daina et al., 2017). In the area of drug discovery and development, initial analysis of a compound's toxicity is critical (C. Li et al., 2019). The Toxicity Estimation Software Tool (T.E.S.T) version 4.2.1 was used to assess the in-silico toxicity of the targeted participants. Using Quantitative Structure-Activity Relationships (QSARs) techniques, the software estimates the toxicity of specified compounds.

## Result and Discussions

### Sequence Retrieval

The uncharacterized protein of the *Photobacterium damselae*, (Amino acid no. 425) which has an unknown function, was obtained from the UniprotKb database under the ID number A0PB24. (Figure 1)

### Similarity Identification and Phylogeny Analysis

The BLASTp result against non-redundant database showed homology with other proteins (Table 1). Multiple sequence alignment ([Supplementary Figure](#)) was considered the FASTA sequences of the uncharacterized protein (A0PB24) and the homologous

MTKVNHLQSLCVI HVDFDIWSGQTRL SASDLKLGEGEI PPEKVAQLGSKKICDPAKLGK  
 FHRLKTETRRLLLKFGMPFMNGFAVPVSKTDEICNKLNDINFQFNQLKQDFIKGYNKAVD  
 EWCQENPEYERAI RAGALPKETVEERIGFEYQVFMIQPVNEDEANAKRLNRKVERLGDDL  
 ISEVVQEANKFYMERLAGRDQCAVTTTRQTLRNIRDKVDGLSFLNSAFNPLVKLLDQTLRG  
 YEQHADGRNIVAPFFYQVVAVLIMSERDRIEQYANGSITVEGMANDIGGSGAQMGRSK  
 DEKAEQKSDKAGELIPATEGGETKQQVGGTESVQSEQTNSGGNAVDLDEDIDNFFKSFA  
 ERGEGESEDES NAGDVVREERVDVEDEPVLPEETPVEQEPVQEEPTEEPLNQELPKTDDD  
 GDYFF

**Figure 1.** Amino acid sequence of uncharacterized protein.

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

Protein ID	Organism	Accession No. AOPB24		
		Protein Name	Identity (%)	e value
WP_001909290.1	<i>Klebsiella</i>	DUF3150 domain-containing protein	100	0.00
WP_000170086.1	<i>Gammaproteobacteria</i>	DUF3150 domain-containing protein	100	0.00
WP_223385441.1	<i>Klebsiella pneumoniae</i>	DUF3150 domain-containing protein	100	0.00
HAD9478497.1	<i>Salmonella enterica</i>	TPA: DUF3150 domain-containing protein	99.76	0.00
AFV41037.1	<i>Aeromonas hydrophila</i>	hypothetical protein	99.25	0.00

annotated proteins. For the confirmation of homology assessment between the proteins, down to the complex and subunit level, phylogenetic analysis was also performed. Because they were discovered in the same class as the query protein, the phylogram allowed us infer that the closely related homologs with the query protein had comparable functions. Phylogenetic tree was constructed based on the alignment and BLAST result which gives the similar concept about the protein (Figure 2). The distances between branches are also included.

#### Conserved Domain, Motif and Fold Identification

The conserved domain search tool revealed that this uncharacterized protein sequence was found to have two domains, DUF3150 (accession No. pfam11348) and 2A1904 super family domain (accession No. cl36772). The result was also checked by two other domain searching tools namely InterProScan and Pfam. Pfam server predicted the Protein of unknown function (DUF3150) at 7–266 amino acid residues with an e-value of  $1.4e^{-89}$ . InterproScan server predicted Protein of unknown function DUF3150 at 7-267 amino acid residues. PF11348, Protein of unknown function (DUF3150), PF10973, Protein of unknown function (DUF2799) and PF20441, Terminase large subunit, endonuclease domain was also found by Motif server. Fold pattern recognition by PFP-FunDSeqE tool revealed the presence of a flavodoxin-like fold within the protein sequence. The flavodoxin fold is a common  $\alpha/\beta$  protein fold, second only to the TIM barrel fold. It has three layers, with two  $\alpha$ -helical layers sandwiching a 5-stranded parallel  $\beta$ -sheet.

#### Physicochemical Characterization

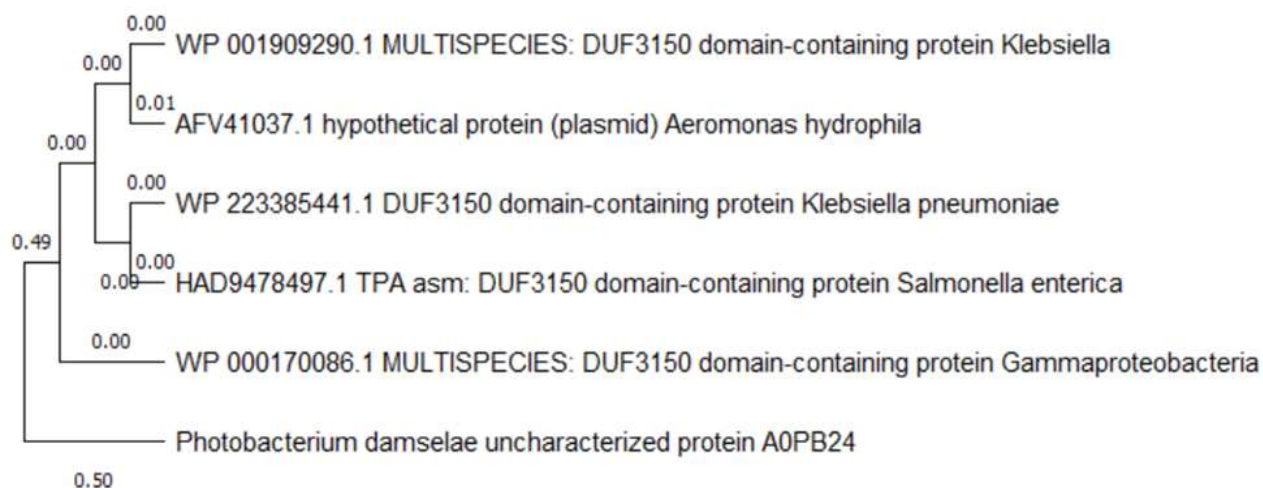
The FASTA format was utilized as a query sequence for physicochemical parameter determination. This protein has an instability index of 44.73 (45), indicating that the protein is stable (Guruprasad et al., 1990). The protein has a molecular weight of 47729.04 kDa and is acidic (pI 4.59), indicating the presence of ribonuclease A residues based on high extinction coefficient values ( $22920 \text{ M}^{-1} \text{ cm}^{-1}$ ) (Gill & von Hippel, 1989). The query protein's higher aliphatic index values (73.39) appear to be a key component in enhanced thermo stability over a wide temperature range. The protein is hydrophilic, and the possibility of better interaction with water (Saikat & Ripon, 2020) as indicated by the lower grand average of hydropathicity (GRAVY) indices value (-0.699).

#### Secondary Structure Inquiry

The alpha helix (Hh), extended strand (Ee), beta turn (Tt), and random coil (Cc) of the protein (AOPB24) were predicted by the SOPMA software to be 218 (51.29%), 27 (6.35%), 22 (5.08%), and 158 (37.18%), accordingly.

#### Subcellular Localization

The subcellular localization of proteins has been used to describe them as therapeutic and vaccination targets (Acharya & Garg, 2016). Proteins found in the cytoplasmic matrix might be used as medication targets, while proteins found in the inner and outer membranes could be used as vaccination targets (Mou et al., 2021;



**Figure 2.** Multiple sequence phylogenetic tree.

Prabhu et al., 2020). The understanding of localization is a crucial step in determining the function of proteins. The protein was anticipated to be present in any of the cell's places (cytoplasm, periplasm, extracellular, inner membrane, and outer membrane) based on the knowledge of trained data sets. Based on the concurrence hits, we estimated the location of the A0PB24 protein subject for the study. The protein's predicted outcome was discovered in the cytoplasm (CELLO score 3.445). The ability to predict proteins is critical for a better understanding of drug targets and the development of effective therapeutic compounds.

### Protein Binding Sites and GO Prediction

Predict protein server was applied for the determination of binding sites prediction where two different protein binding sites (reliability score: 35) were identified at positions 224;253-255. and 4 different polynucleotide binding sites (DNA binding; reliability score: 53) were identified at position 19-29; 48-68; 77-86; 199-214; and different polynucleotide binding sites (RNA binding; reliability score: 42) were identified 51-55;57-63;65;201-210. (Figure 3).

The functional aspects of biological process ontology, molecular functional ontology and cellular component ontology were predicted and categorized using gene ontology (Table 5). Molecular Function Ontology calculated as Protein binding (43%); Biological process ontology (Table 2) detected as Viral RNA genome packing (43%); Multi-organism process (43%). Cellular component ontology (Table 4) predicted as helical viral capsid (43%); host cell cytoplasm (43%); viral nucleocapsid (43%); virion (43%).

### Three-Dimensional Structure Prediction

The HHPred service was used for tertiary modeling, and the modeled protein had a sequence identity of 88.8 %. At the Modeller server, the homology derived model

was produced from HHPred using 3RIU\_A as the template and the tertiary modeled protein structure was saved in PDB format predicted by Modeller. It was estimated using the relevance of threading template alignments and the structure assembly simulations' convergence parameters, then visualized using the Discovery Studio.

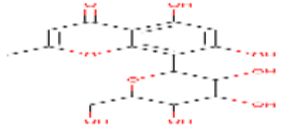
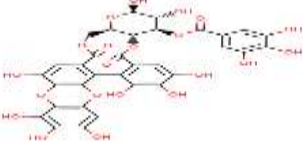
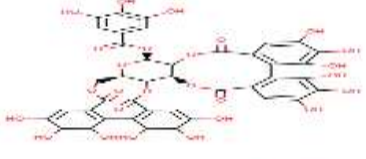
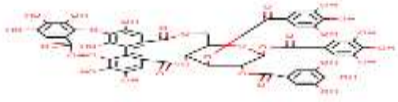
### 3D Structure Refinement

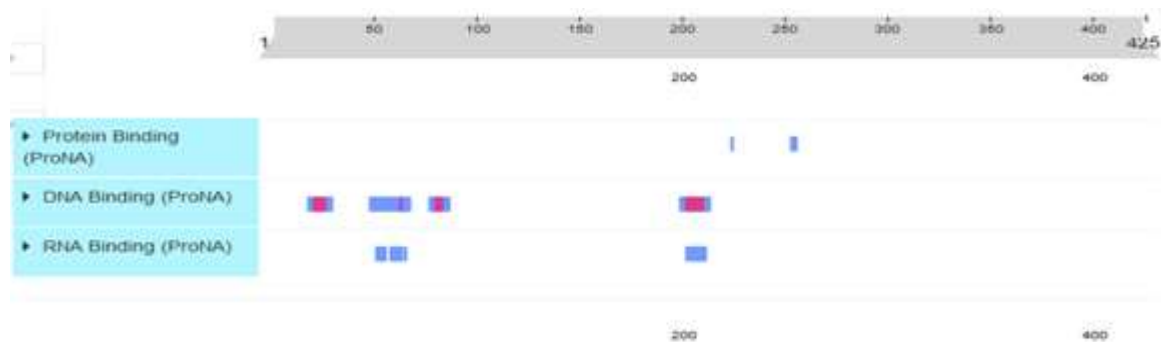
The Galaxy Refine server was used to refine the protein's projected tertiary structure, resulting in five refined models with more amino acid residues in the preferred location. The scores shown above indicate the enhanced model's caliber as compared to the other variants. In Discovery Studio, refine model 1 was chosen and visualized (Figure 4).

### Model Quality Assessment

The before and after improved protein models (3RIU\_A) were validated using the Ramachandran Plot Server and the ProSA-Web server. According to the Ramachandran plot server, 97.024 % of the structure before refinement was in the favorable zone followed by 1.786% and 1.190% in preferred region and questionable region, respectively. The rampage server generated a better result after refinement, with 98.656% of residues in highly preferred regions whereas 1.075 % and 0.269 % in preferred region and questionable region, respectively which indicate that this model quality is significant. The validation quality and potential faults in a basic tertiary structure model are assessed using the ProSA-web server. Validation of the final 3RIU\_A protein model reveals a Z-score of -5.57 which indicate the model is significant. The tertiary structure of the projected two-chain protein model (3RIU\_A) passed the Verify 3D structure assessment experiment with ease. From 141 to 165 amino acids, good quality was reported in VERIFY3D.

**Table 2.** *Syzygium aromaticum* molecular docking score and ligand structure

Ligand compound name and ID	Binding Affinity (kcal/mol)	Ligand Structure
1,5-Anhydro-1-(5,7-dihydroxy-2-methyl-4-oxo-4H-chromen-8-yl)hexitol CHEMSPIDER:35014984	-7.4	
Syzyginin B CID:102445430	-8.1	
Casuarictin CID:118701062	-7.8	
Syzyginin A HMDBID: HMDB30236	-7.6	

**Figure 3.** Protein-Protein and polynucleotide binding sites**Figure 4.** 3D structures of crude model (A) and refine model (B) of the protein

### Protein-protein Interaction

The STRING protein-protein interaction network indicated that our putative protein interacts strongly with the mixed, incl. Type-IV secretion system protein TraC, and Type-F conjugative transfer system pilin assembly protein (PSNIH2\_20955, 1484157.PSNIH2\_20965, 1484157.PSNIH2\_20970, 1484157.PSNIH2\_20980, 1484157.PSNIH2\_2098); this connection suggests that the protein may serve as a DNA/RNA hydrazase. (Figure 5)

### Active Site Detection

The modeled protein's CASTp v.3.0 algorithm predicted 14 distinct active sites (Figure 6). CASTp is a database server that can detect areas on proteins, define their boundary, compute the area, and find the areas' dimensions. Pockets on protein surfaces and vacuums hidden within proteins are involved. Surfaces of solvent accessible molecules (surface of Richards) and molecular surfaces (surface of Connolly) are used to define a pocket and volume spectrum or vacuum. CASTp might be used to investigate protein operational zones and surface characteristics. CASTp provides a graphical, user-interface flexible, dynamic display as well as on-the-fly measurement of user-submitted constructions

(Tian et al., 2018). The top active sites of the modeled protein were identified between the area of 298.358 and the volume of 207.309. Figure 7 shows the protein's anticipated active site together with its amino acid residues.

### Retrieval and Preparation of Phytochemicals and Conventional Drugs

The phytochemical components found in *S. aromaticum* were extracted and recorded in a 2D (SDF) file format from IMPPAT. Drug Bank database was used to retrieve the drug compounds in SDF file format. During the ligand preparation procedures, the compounds were produced and optimized, then converted to pdbqt file format for further assessment.

### Molecular Docking Analysis

A molecular docking study was first conducted to screen and identify the optimal intermolecular interaction among the desired protein and phytochemical substances. PyRx tools AutoDock Vina wizard was used to perform molecular docking between selected four phytochemical compounds and the protein of choice. (Table 3)

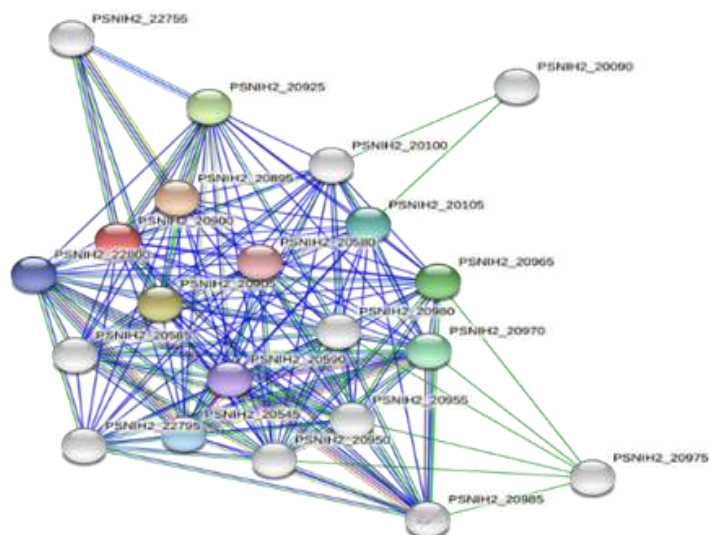
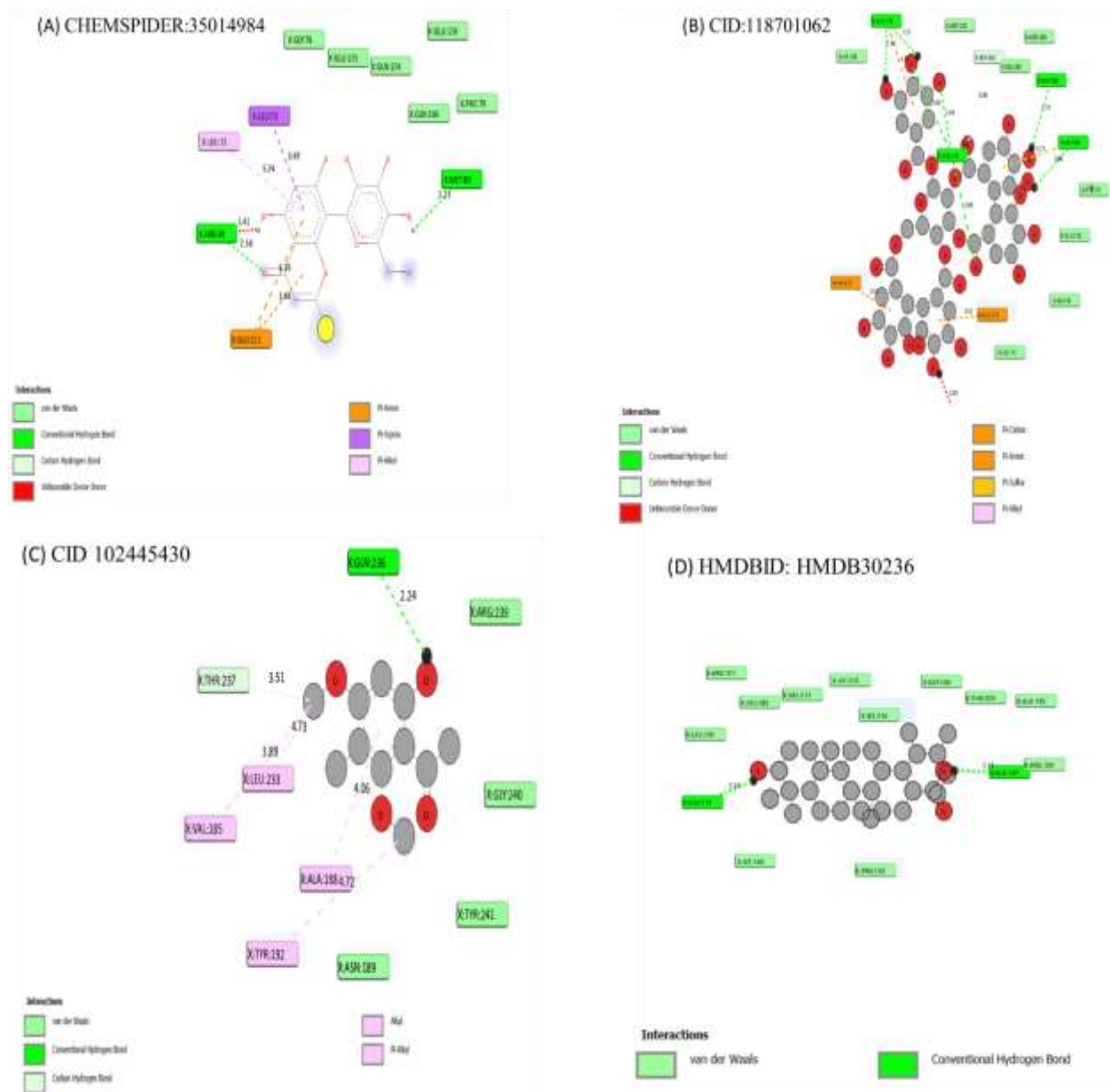


Figure 5. String network analysis of the uncharacterized protein



Figure 6. The uncharacterized protein's active location. (A) The blue sphere represents the protein's active site. (B) The active site of amino acid residues (Blue color).





**Figure 7.** Depicted the 2D interaction between the selected phytochemical compounds and 3RIU\_A protein. (A): The interaction between the uncharacterized protein and CHEMSPIDER:35014984. (B): The interaction between the uncharacterized protein and CID CID:118701062 compound. (C): The interaction between the uncharacterized protein and CID 102445430 compound. (D): The interaction between the uncharacterized protein and HMDBID: HMDB30236 compound.

**Table 3.** Selected Drugs molecular docking score and ligand structure

Ligand	Binding Affinity (kcal/mol)	Ligand Structure
Flumequine DB: 08972	-5.9	
Oxytetracycline DB: 00595	-6.8	
Florfenicol DB: 11413	-5.7	

### Interpretations of Protein–ligands Interaction

With the desired protein (UniprotKb ID: AOPB24), the compound CHEMSPIDER:35014984 produced two conventional hydrogen bond interactions with MET180 and ARG69, where one hydrophobic pi-Alkyl bonds were discovered to form at the positions LEU72 in Figure 7A and Table 4.

The interaction investigation of the compound CID:102445430 revealed two hydrogen bonds at the location of GLN236 (2.24) and THR237 (3.51) Figure 7C and Table 4.

With the target protein, compound CID:118701062 has been found to create a conventional hydrogen bonds at the positions of ASP178 (2.17), MET80 (3.06), GLN186 (2.79), GLN124 (2.45) and SER182 (3.38), where one Pi-Anion and Pi-Cation bonds each have been noted at the positions of GLU121 (3.52) and LYS117 (3.13) Figure 7B and Table 4. Carbon hydrogen bond bonds was observed to form exclusively at the GLU174 (2.13) position of the molecule HMDBID: HMDB30236as shown in Figure 7D and Table 4.

On the other hand, conventional chemical drug DB:00595 has been discovered to make multiple conventional hydrogen bonds with the target protein model (UniprotKb ID: AOPB24). Three conventional hydrogen bonds have been discovered to form at the sites of THR209, ASN212, and THR205 whereas one carbon hydrogen bond form at the sites of ALA135 (Figure 8B; Table.5). However, only one carbon–hydrogen bond was detected during the interaction of

the conventional drug DB: 08972 with GLY177 and DB: 11413 with LYS140 (Figure 8A and Table 5).

In this study, we identified 4 potential drug compounds from *S. aromaticum* by molecular docking and other process. Initially, molecular docking process has used to screen the compounds, where the 3 out of 4 phytochemical compounds has been with the highest binding affinities compare to conventional drug. These three compounds (CHEMSPIDER:35014984, CID:102445430 and CID:118701062) also found with more hydrogen bond binding sites with the uncharacterized protein. For drug target or any potential antibacterial inhibitor prediction, hydrogen bond interaction is pivotal for not only in drug-target interactions but also in holding together the structure of proteins and DNA (Wade & Goodford, 1989). However, conventional antibacterial drug found less hydrogen bond interactions with the modeled uncharacterized protein and the docking score also lesser than the phytochemical compounds which indicate that less effective against this protein. However, these can be chosen for further validation through in-vivo experiment.

### Absorption, Distribution, Metabolism, and Excretion (ADME) and Toxicity Test

One of the most important characteristics for designing successful drug candidates against a certain target is ADMET properties. The Lipinski rule of five (RO5) was used to access drug similarity features of

**Table 4.** The following is a list of bonding interactions between four phytochemicals and the uncharacterized protein

Ligand ID	Distance	Category	Types	Binding Site
CHEMSPIDER:35014984	2.21;2.58	Hydrogen Bond	Conventional Hydrogen Bond	MET180; ARG69
	3.68	Electrostatic	Pi-Anion	GLU121
	5.24	Hydrophobic	Pi-Alkyl	LEU72
CID:102445430	2.24	Hydrogen Bond	Conventional Hydrogen Bond	GLN236
	3.51	Hydrogen Bond	Carbon Hydrogen Bond	THR237
	4.72	Hydrophobic	Pi-Alkyl	TYR192
	4.72;4.73	Hydrophobic	Alkyl	ALA188; LEU233
CID:118701062	2.17;3.06	Hydrogen Bond	Conventional Hydrogen Bond	ASP178; MET80
	2.79;2.45	Hydrogen Bond	Conventional Hydrogen Bond	GLN186; GLN124
	3.38	Hydrogen Bond	Carbon hydrogen bond	SER182
	3.13	Electrostatic	Pi-Cation	LYS117
	3.52	Electrostatic	Pi-Anion	GLU121
HMDBID: HMDB30236	2.13	Conventional Hydrogen Bond	Carbon hydrogen bond	GLU174

**Table 5.** The following is a list of bonding interactions between 3 conventional drug and the uncharacterized protein

Ligand ID	Distance	Category	Types	Binding Site
DB: 08972	3.58	Hydrogen bond	Carbon Hydrogen Bond	GLY177
	4.92	Hydrophobic	Alkyl	LYS216
DB: 00595	2.92;2.22;2.73	Hydrogen bond	Conventional Hydrogen Bond	THR209; ASN212; THR205
	4.02	Electrostatic	Pi-Cation	LYS216
	3.56	Hydrogen bond	Carbon hydrogen bond	ALA135
DB: 11413	2.27	Hydrogen bond	Conventional Hydrogen Bond	LYS140
	4.86; 4.32	Electrostatic	Pi-Cation	GLU174; LYS216
	5.11	Hydrophobic	Pi Alkyl	ILE213



through different lab-based experiment techniques can help to determine the activity of the compound that will provide alternatives for disease immunotherapy.

### Ethical Statement

Not Applicable

### Funding Information

There is no any funding institution for this study.

### Author Contribution

Conceptualization, M.M. and SK; methodology, SK. and M.M.; software, SK.; validation, M.M., and SK.; formal analysis, SK.; investigation, SK.; resources, SK.; data curation, M.M. and SK.; writing— original draft preparation, Sk.; writing—review and editing, SK. 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

- Abushouk, A. I., Negida, A., Ahmed, H., & Abdel-Daim, M. M. (2017). Neuroprotective mechanisms of plant extracts against MPTP induced neurotoxicity: Future applications in Parkinson's disease. *Biomedicine & Pharmacotherapy*, 85, 635-645. <https://doi.org/https://doi.org/10.1016/j.biopha.2016.11.074>
- Acharya, A., & Garg, L. C. (2016). Drug Target Identification and Prioritization for Treatment of Ovine Foot Rot: An In Silico Approach. *International journal of genomics*, 2016, 7361361-7361361. <https://doi.org/10.1155/2016/7361361>
- Andreoni, F., & Magnani, M. (2014). Photobacteriosis: prevention and diagnosis. *Journal of immunology Research*, 2014.
- Alzohairy, A. (2011). BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences*, 2, 60-61.
- Balado, M., Benzekri, H., Labella, A. M., Claros, M. G., Machado, M., Borrego, J. J., Osorio, C. R., & Lemos, M. L. (2017). Genomic analysis of the marine fish pathogen *Photobacterium damsela* subsp. *piscicida*: Insertion sequences proliferation is associated with chromosomal reorganisations and rampant gene decay. *Infection*, *Genetics and Evolution*, 54, 221-229. <https://doi.org/https://doi.org/10.1016/j.meegid.2017.07.007>
- Batiha, G. E.-S., Alkazmi, L. M., Wasef, L. G., Beshbishy, A. M., Nadwa, E. H., & Rashwan, E. K. (2020). *Syzygium aromaticum* L. (Myrtaceae): Traditional Uses, Bioactive Chemical Constituents, Pharmacological and Toxicological Activities. *Biomolecules*, 10(2), 202. <https://doi.org/10.3390/biom10020202>
- Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*, 2(9), 1511-1519. <https://doi.org/10.1002/pro.5560020916>
- Combet, C., Blanchet, C., Geourjon, C., & Deléage, G. (2000). NPS@: Network Protein Sequence Analysis. *Trends in Biochemical Sciences*, 25(3), 147-150. [https://doi.org/10.1016/S0968-0004\(99\)01540-6](https://doi.org/10.1016/S0968-0004(99)01540-6)
- Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*, 7, 42717. <https://doi.org/10.1038/srep42717>
- Dallakyan, S., & Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. *Methods Mol Biol*, 1263, 243-250. [https://doi.org/10.1007/978-1-4939-2269-7\\_19](https://doi.org/10.1007/978-1-4939-2269-7_19)
- 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>
- Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol*, 277, 396-404. [https://doi.org/10.1016/s0076-6879\(97\)77022-8](https://doi.org/10.1016/s0076-6879(97)77022-8)
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., & Jensen, L. J. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*, 41(Database issue), D808-815. <https://doi.org/10.1093/nar/gks1094>
- Gazi, M. A., Kibria, M. G., Mahfuz, M., Islam, M. R., Ghosh, P., Afsar, M. N., Khan, M. A., & Ahmed, T. (2016). Functional, structural and epitopic prediction of hypothetical proteins of *Mycobacterium tuberculosis* H37Rv: An in silico approach for prioritizing the targets. *Gene*, 591(2), 442-455. <https://doi.org/10.1016/j.gene.2016.06.057>
- Hirokawa, T., Boon-Chieng, S., & Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, 14(4), 378-379. <https://doi.org/10.1093/bioinformatics/14.4.378>
- Hughes, J. P., Rees, S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. *British journal of pharmacology*, 162(6), 1239-1249. <https://doi.org/10.1111/j.1476-5381.2010.01127.x>
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283-291. <https://doi.org/https://doi.org/10.1107/S0021889892009944>
- Li, C., Wang, J., Wang, Y., Gao, H., Wei, G., Huang, Y., Yu, H., Gan, Y., Wang, Y., Mei, L., Chen, H., Hu, H., Zhang, Z., & Jin, Y. (2019). Recent progress in drug delivery. *Acta Pharm Sin B*, 9(6), 1145-1162.

- <https://doi.org/10.1016/j.apsb.2019.08.003>
- Li, Y., Meng, Q., Yang, M., Liu, D., Hou, X., Tang, L., Wang, X., Lyu, Y., Chen, X., Liu, K., Yu, A. M., Zuo, Z., & Bi, H. (2019). Current trends in drug metabolism and pharmacokinetics. *Acta Pharm Sin B*, 9(6), 1113-1144. <https://doi.org/10.1016/j.apsb.2019.10.001>
- Liang, P. H., Cheng, W. C., Lee, Y. L., Yu, H. P., Wu, Y. T., Lin, Y. L., & Wong, C. H. (2006). Novel five-membered iminocyclitol derivatives as selective and potent glycosidase inhibitors: new structures for antivirals and osteoarthritis. *Chembiochem*, 7(1), 165-173. <https://doi.org/10.1002/cbic.200500321>
- Lim, S. M., Xie, T., Westover, K. D., Ficarro, S. B., Tae, H. S., Gurbani, D., Sim, T., Marto, J. A., Jänne, P. A., Crews, C. M., & Gray, N. S. (2015). Development of small molecules targeting the pseudokinase Her3. *Bioorganic & Medicinal Chemistry Letters*, 25(16), 3382-3389. <https://doi.org/https://doi.org/10.1016/j.bmcl.2015.04.103>
- Lubec, G., Afjehi-Sadat, L., Yang, J.-W., & John, J. P. P. (2005). Searching for hypothetical proteins: Theory and practice based on original data and literature. *Progress in Neurobiology*, 77(1), 90-127. <https://doi.org/https://doi.org/10.1016/j.pneurobio.2005.10.001>
- Madden, T. L., Busby, B., & Ye, J. (2019). Reply to the paper: Misunderstood parameters of NCBI BLAST impacts the correctness of bioinformatics workflows. *Bioinformatics*, 35(15), 2699-2700. <https://doi.org/10.1093/bioinformatics/bty1026>
- Magariños, B., Romalde, J. L., Bandín, I., Fouz, B., & Toranzo, A. E. (1992). Phenotypic, antigenic, and molecular characterization of *Pasteurella piscicida* strains isolated from fish. *Appl Environ Microbiol*, 58(10), 3316-3322. <https://doi.org/10.1128/aem.58.10.3316-3322.1992>
- Marchler-Bauer, A., & Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res*, 32(Web Server issue), W327-331. <https://doi.org/10.1093/nar/gkh454>
- Mladineo, I., Miletić, I., & Bočina, I. (2006). Photobacterium damsela subsp. piscicida Outbreak in Cage-Reared Atlantic Bluefin Tuna *Thunnus thynnus*. *Journal of Aquatic Animal Health*, 18(1), 51-54. <https://doi.org/https://doi.org/10.1577/H05-012.1>
- Mohanraj, K., Karthikeyan, B. S., Vivek-Ananth, R. P., Chand, R. P. B., Aparna, S. R., Mangalapandi, P., & Samal, A. (2018). IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics. *Scientific Reports*, 8(1), 4329. <https://doi.org/10.1038/s41598-018-22631-z>
- Mou, 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>
- Pollastri, M. P. (2010). Overview on the Rule of Five. *Current Protocols in Pharmacology*, 49(1), 9.12.11-19.12.18. <https://doi.org/https://doi.org/10.1002/0471141755.p0912s49>
- Prabhu, D., Rajamanikandan, S., Anusha, S. B., Chowdary, M. S., Veerapandian, M., & Jeyakanthan, J. (2020). In silico Functional Annotation and Characterization of Hypothetical Proteins from *Serratia marcescens* FGI94. *Biology Bulletin*, 47(4), 319-331. <https://doi.org/10.1134/S1062359020300019>
- Saikat, A. S. M., & Ripon, A. (2020). Structure Prediction, Characterization, and Functional Annotation of Uncharacterized Protein BCRIVMBC126\_02492 of *Bacillus cereus*: An In Silico Approach. *American Journal of Pure and Applied Biosciences*, 2, 104-111. <https://doi.org/10.34104/ajpab.020.01040111>
- Seeliger, D., & de Groot, B. L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *Journal of computer-aided molecular design*, 24(5), 417-422. <https://doi.org/10.1007/s10822-010-9352-6>
- Shen, H. B., & Chou, K. C. (2007). Virus-PLoc: a fusion classifier for predicting the subcellular localization of viral proteins within host and virus-infected cells. *Biopolymers*, 85(3), 233-240. <https://doi.org/10.1002/bip.20640>
- Shieh, W. Y., Chen, Y. W., Chaw, S. M., & Chiu, H. H. (2003). *Vibrio ruber* sp. nov., a red, facultatively anaerobic, marine bacterium isolated from sea water [Review]. *International Journal of Systematic and Evolutionary Microbiology*, 53(2), 479-484. <https://doi.org/10.1099/ijs.0.02307-0>
- Söding, J., Remmert, M., Biegert, A., & Lupas, A. N. (2006). HHSenser: exhaustive transitive profile search using HMM-HMM comparison. *Nucleic Acids Research*, 34(Web Server issue), W374-W378. <https://doi.org/10.1093/nar/gkl195>
- Szymański, P., Markowicz, M., & Mikiciuk-Olasik, E. (2012). Adaptation of high-throughput screening in drug discovery-toxicological screening tests. *Int J Mol Sci*, 13(1), 427-452. <https://doi.org/10.3390/ijms13010427>
- 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. <https://doi.org/10.1093/nar/gky473>
- Toranzo, A. E., & Barja, J. L. (1993). Virulence factors of bacteria pathogenic for coldwater fish. *Annual Review of Fish Diseases*, 3, 5-36. [https://doi.org/https://doi.org/10.1016/0959-8030\(93\)90027-9](https://doi.org/https://doi.org/10.1016/0959-8030(93)90027-9)
- Urbanczyk, H., Ast, J. C., & Dunlap, P. V. (2011). Phylogeny, genomics, and symbiosis of *Photobacterium*. *FEMS Microbiology Reviews*, 35(2), 324-342. <https://doi.org/10.1111/j.1574-6976.2010.00250.x>
- Wade, R. C., & Goodford, P. J. (1989). The role of hydrogen-bonds in drug binding. *Prog Clin Biol Res*, 289, 433-444.
- Wichapong, K., Nueangaudom, A., Pianwanit, S., Sippl, W., & Kokpol, S. (2013). Identification of potential hit compounds for Dengue virus NS2B/NS3 protease inhibitors by combining virtual screening and binding free energy calculations. *Trop Biomed*, 30(3), 388-408.
- 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), W407-410. <https://doi.org/10.1093/nar/gkm290>
- Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*, 112, 531-552. <https://doi.org/10.1385/1-59259-584-7:531>
- Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., Wang, Q., Xu, Y., Li, M., Li, X., Zheng, M., Chen, L., & Li, H. (2020). Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm Sin B*, 10(5), 766-788. <https://doi.org/10.1016/j.apsb.2020.02.008>

- Wu, J., Hu, B., Sun, X., Wang, H., Huang, Y., Zhang, Y., Liu, M., Liu, Y., Zhao, Y., Wang, J., & Yu, Z. (2020). In silico study reveals existing drugs as  $\alpha$ -glucosidase inhibitors: Structure-based virtual screening validated by experimental investigation. *Journal of Molecular Structure*, 1218, 128532. <https://doi.org/10.1016/j.molstruc.2020.128532>
- Zhang, X.-H., He, X., & Austin, B. (2020). *Vibrio harveyi*: a serious pathogen of fish and invertebrates in mariculture. *Marine life science & technology*, 1-15. <https://doi.org/10.1007/s42995-020-00037-z>