Functional Analysis of Hypothetical Proteins of *Vibrio parahaemolyticus* Reveals the Presence of Virulence Factors and Growth-Related Enzymes With Therapeutic Potential

Sazzad Shahrear[1], Maliha Afroj Zinnia[2], Md. Rabi Us Sany[1] and Abul Bashar Mir Md. Khademul Islam[1]

[1]Department of Genetic Engineering and Biotechnology, University of Dhaka, Dhaka, Bangladesh. [2]Department of Pharmacy, East West University, Dhaka, Bangladesh.

ABSTRACT: *Vibrio parahaemolyticus*, an aquatic pathogen, is a major concern in the shrimp aquaculture industry. Several strains of this pathogen are responsible for causing acute hepatopancreatic necrosis disease as well as other serious illness, both of which result in severe economic losses. The genome sequence of two pathogenic strains of *V. parahaemolyticus*, MSR16 and MSR17, isolated from Bangladesh, have been reported to gain a better understanding of their diversity and virulence. However, the prevalence of hypothetical proteins (HPs) makes it challenging to obtain a comprehensive understanding of the pathogenesis of *V. parahaemolyticus*. The aim of the present study is to provide a functional annotation of the HPs to elucidate their role in pathogenesis employing several in silico tools. The exploration of protein domains and families, similarity searches against proteins with known function, gene ontology enrichment, along with protein-protein interaction analysis of the HPs led to the functional assignment with a high level of confidence for 656 proteins out of a pool of 2631 proteins. The in silico approach used in this study was important for accurately assigning function to HPs and inferring interactions with proteins with previously described functions. The HPs with function predicted were categorized into various groups such as enzymes involved in small-compound biosynthesis pathway, iron binding proteins, antibiotics resistance proteins, and other proteins. Several proteins with potential druggability were identified among them. In addition, the HPs were investigated in search of virulent factors, which led to the identification of proteins that have the potential to be exploited as vaccine candidate. The findings of the study will be effective in gaining a better understanding of the molecular mechanisms of bacterial pathogenesis. They may also provide an insight into the process of evaluating promising targets for the development of drugs and vaccines against *V. parahaemolyticus*.

KEYWORDS: Functional annotation, hypothetical protein, virulence factor, biosynthesis, therapeutic potential

Introduction

Litopenaeus vannamei and *Penaeus monodon* are two of the most economically important species of farmed shrimp grown in Asia. Unfortunately, the emergence of various viral, bacterial, and fungal infections continues to wreak havoc on shrimp productivity.[1-3] Vibriosis is one of the most prevalent illnesses in Asia, causing severe morbidity in farmed aquatic products (ie, shrimp, fish, and shellfish).[4] *Vibrio harveyi*, *V. anguillarum*, *V. alginolyticus* and *V. parahaemolyticus* are examples of opportunistic *Vibrio* pathogens with particularly virulent strains.[5] In 2009, acute hepatopancreatic necrosis disease (AHPND) had a quick, devastating effect on early stages of shrimp during an initial outbreak elsewhere in southwest of China.[6] Since then, it has expanded all over the world, posing a severe threat to the shrimp industry in several Asian countries.[7]

When shrimp are 1 month old or post larvae are approximately 20 to 30 days old, AHPND can potentially result up to 100% morbidity.[8] *V. parahaemolyticus* infection provokes an atrophied, pale hepatopancreas, severe sloughing off of the hepatopancreatic epithelial cells, destruction of the brush boundary of the anterior midgut of the shrimp and hemocyte infiltration.[1,8-10] *V. parahaemolyticus* is a facultative anaerobic, gram-negative bacterium that could be found in estuarine, marine, and coastal ecosystems across the world, including in shrimp aquaculture.[11,12] Ever since discovery, it has been associated to human gastrointestinal illness, septicemia, and wound infections and is considered a causative agent of foodborne illness over the world.[12-14]

Only the strains of opportunistic marine pathogen *V. parahaemolyticus* (named as VP_{AHPND}) carrying a 69-kbp virulent pVA1 plasmid encoding binary toxin genes homologous to the *Photorhabdus* insect-related (Pir) toxin, PirA and PirB, were found to produce AHPND.[6,15,16] VP_{AHPND} lacks its potential for causing AHPND when the pVA1 plasmid has been removed or the Pir genes are knocked down selectively.[16] The plasmid also contains conjugative transfer genes as well as transposons, implying that the plasmid might be mobilized into some other strains or even into other species.[1,16] Interestingly, after being reported in *V. parahaemolyticus*, the pVA1 plasmid was also reported in a variety of *Vibrio* species, including *V. campbelli*, *V. owensii*, *V. punensis*, and *V. harveyi*, all of which have the potential to cause AHPND.[1] Even though

the plasmid-encoded binary toxins pirA and pirB have been identified as the leading cause of AHPND in shrimp, additional virulence factors reported in *V. parahaemolyticus* might play a key role during infection.[17,18]

The shrimp production has been severely impacted by the AHPND causing strain of *V. parahaemolyticus*, VP$_{AHPND}$.[19] The bacteria are transmitted orally and subsequently accumulates inside the gastrointestinal tract of the shrimp, where it produces and secretes binary toxins (ie, PirA^{VP} and PirB^{VP}) which induces tissue destruction and invalidism of the hepatopancreas of the shrimp digestive system.[16,20] Farmers sought to prevent AHPND epidemics by eradicating and renovating their ponds, but they have been unable to stop outbreaks once the infection emerged repeatedly in farmlands.[5] To eliminate AHPND, the etiologic agent, *V. parahaemolyticus*, has to be extensively characterized.

Two isolates of *V. parahaemolyticus* (strains MSR16 and MSR17) were isolated from cultured shrimp (*P. monodon*) in the southwest region of Bangladesh, and their genomes had been sequenced and published, way that allows genomics and proteomics analyses to gain a detailed understanding of microbial virulence mechanisms and underlying pathogenesis.[18] Nonetheless, gaining a better understanding of these mechanisms remains a challenge. During proteomic studies, functional annotation is essential to determine the function of proteins.[21] In the meanwhile, the function of a large number of coding sequences still remains unknown. Pathogenesis and virulence determination have been challenging to fully understand due to a lack of comprehensive proteome data due to coding sequences without a proper prediction of functions. The term "hypothetical protein" (HP) is used to describe these molecules. The majority of such proteins are thought to play a significant role to play in the cell; therefore, proper annotation can potentially lead to new insights into their structures, functions, and pathways.[22,23]

Considering wet lab approaches are often time-consuming, expensive, and labor intensive for unraveling functions of desired proteins, in silico approaches have emerged as significant methods for predicting or identifying the functions of hypothetical proteins. Due to the obvious similarity with known proteins, homology-based functional annotation may be used to assign functions to HPs.[21,24-26] Bioinformatics approaches, particularly protein-protein interactions, can aid in the proper characterization of the biological processes in which HPs are involved.[27] The objective of this study was to assign functions to the hypothetical proteins encoded within the genomes of the *V. parahaemolyticus* isolates to identify novel proteins that might aid to better understand the pathogenesis and virulence mechanism of the bacteria as well as identify new therapeutic targets employing computational approaches.

Materials and Methods

The overview of methodology is illustrated in Figure 1.

Retrieval of protein sequence data

The genomes of two isolates of *V. parahaemolyticus* (strains MSR16 and MSR17) isolated from cultured shrimp (*P. monodon*) in the southwest region of Bangladesh were retrieved and analyzed in this study (accession numbers: RPDA00000000.1 and RPDB00000000.1, respectively) from the National Center for Biotechnology Information (NCBI).[28] The strain MSR16 genome encoded a total of 5479 genes, 1403 of which were annotated as HP, whereas the strain MSR17 genome encoded a total of 5187 genes, 1228 of which were annotated as HP.[18] Using an in-house python script, the coding sequences (CDS) annotated as hypothetical proteins were retrieved from those genomes (Supplementary Table 1).

Functional annotation of hypothetical proteins

Gene ontology prediction. To annotate the function of the HPs, the protein sequences were initially analyzed using the GO FEAT[29] and PANNZER[30] tools to obtain a preliminary gene ontology (GO) prediction with an *e-value* of 1e-03. For functional annotation of protein-coding genes, the GO vocabulary is used.[31] Protein sequences with GO IDs predicted by both servers were carefully chosen, and domain and function were further investigated with a number of bioinformatics tools.

Family and domain prediction. Multiple databases were scanned for identification of conserved domains to predict protein function based on domain structure. HP sequences were initially evaluated using the HMMER tool, which offers fast screening against frequently used sequence databases, employs profile hidden Markov model libraries for functional annotation of the HP sequences as well as protein families and domains and enables protein homology search algorithms within the HMMER 3.3.2 software suite.[32] For significant *e-values*, the cutoff was set at 0.01. The Pfam[33] and Superfamily[34] databases were used to identify protein families, while the Gene3D[35] database was used to identify protein domains. The InterProScan[36] tool, which scans the InterPro[37] database for matches, was then used to perform functional analysis of proteins by categorizing them into families as well as identifying domains and essential sites encoded by the HPs. The NCBI Batch CD-Search tool was used to compare query HP sequences to databases of conserved domain models using RPS-BLAST.[38] The CDD – 58235 PSSMs database was searched with the NCBI Batch CD-Search tool, with a threshold of 0.01.

HP functions have been predicted up to this point based on domains and families identified by screening databases like as SUPERFAMILY, Pfam, Gene3D, InterPro, and CDD – 58235 PSSMs. InteractiVenn[39] was used to identify HPs with predicted functions from three or more tools.

Finally, the annotated homologous proteins from related organisms were identified using the Basic Local Alignment Search Tool (BLAST).[40] The non-redundant (nr) database of
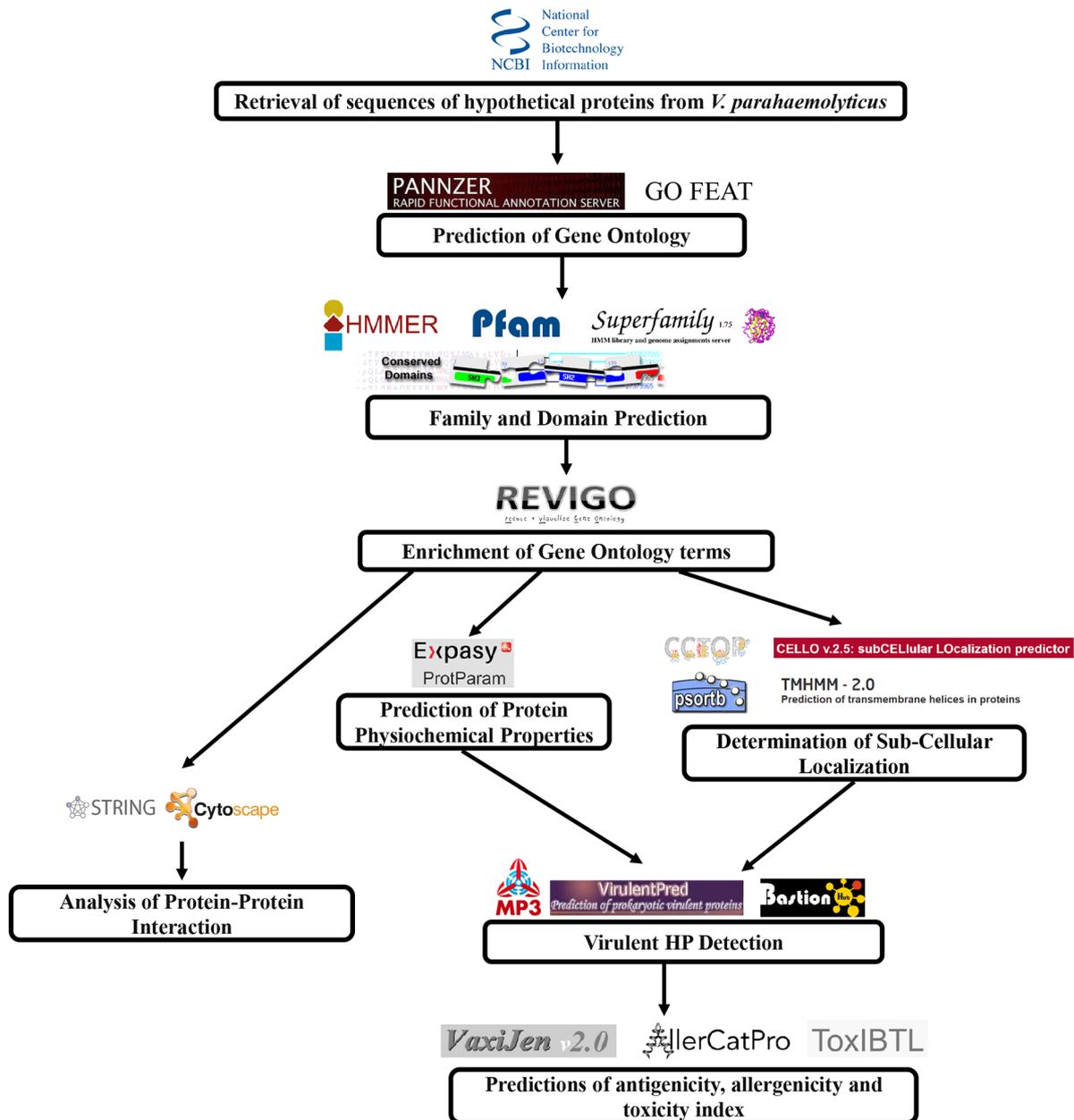
Figure 1. Workflow for functional annotation of *V. parahaemolyticus* hypothetical proteins.

the NCBI was searched for homologs with an identity of more than 90% and an *e-value* of less than 1e-03.[21,41,42] REVIGO[43] was implemented to enrich the GO terms of the annotated HPs and visualize data.

Prediction of protein physiochemical properties. Physical and chemical properties of the virulent HPs were determined using Expasy's ProtParam[44] tool in Linux operating system implemented in Julia (version 1.5.1)[45] with package BioSequences (v2.0.5).

Determination of subcellular localization. PSORTb v3.0[46] and CELLO v.2.5[47] were used in the study to explore the subcellular locations of the HPs using default parameters for gramnegative bacteria. The presence of transmembrane helices as well as the topology of the HPs were predicted using TMHMM 2.0[48] and CCTOP[49] with default parameters. SignalP 6.0,[50] which employs a neural network design using a conditional random field, was used in predicting the presence of signal peptides specific for the secretory (Sec) and the twin-arginine translocation (Tat) pathways, as well as the location of signal peptide cleavage sites.

Virulent HP detection

To identify virulence factors from annotated HPs, they were initially evaluated with the MP3 tool,[51] which uses an SVM and HMM approach to accurately estimate virulent proteins present within genomic and metagenomic data. VirulentPred,[52] another SVM-based virulence prediction tool, was also used to

classify pathogenic proteins in bacteria. Furthermore, using the BastionX prediction method, BastionHub[53] was used to predict substrates for several secretion systems found in gram-negative bacteria (System I-IV, VI). The HPs that were identified to be virulent by all three tools were then identified and further studied.

Predictions of antigenicity, allergenicity, and toxicity index

VaxiJen v2.0[54] server and ANTIGENpro server[55] were used to predict the antigenicity of the vaccine peptide. The ToxIBTL server[56] and the AllerCatPro v. 2.0[57] server were used to predict the peptide's toxicity and allergenicity, respectively.

Analysis of Protein–Protein Interaction (PPI)

Finally, the PPIs for the hypothetical proteins from the *V. parahaemolyticus* MSR16 and MSR17 strains were constructed using the String 11.5 database to validate the functions of the annotated HPs and their interactions with other proteins.[58] Only interactions having score values exceeding 0.700 (high confidence) as well as high FDR stringency (1%) were used to provide the most accurate and reliable PPIs.[21] The *P* value for PPI enrichment was $<10^{-16}$.

In the String 11.5 search, the *V. parahaemolyticus* RIMD 2210633 strain was selected here as the strain with highest similarity. The interolog mapping approach was used to translate the identified interactions to *V. parahaemolyticus* MSR16 and MSR17 strains, which implies that when two proteins interact, their orthologous partners would also interact.[59-61] STRING, as described in eggNOG, exploits hierarchically organized orthologous group relations to retrieve association between relevant species.[62-64]

To perform a more in-depth analysis and understanding of the interactions of the potentially virulent HPs with other proteins and among themselves, Cytoscape 3.9.0[65] was employed. PPI networks were validated using the network analyzer plugin[66] in the Cytoscape 3.9.0 program. Protein molecules were allocated to nodes in Cytoscape, whereas molecular interactions were given to edges.

Results

Function annotation of hypothetical protein to both strains

All of the protein sequences were evaluated employing the GO FEAT and PANNZER tools, which facilitated the prediction of GO annotation. The MSR16 strain of *V. parahaemolyticus* produced a total of 1403 HPs, and GO terms were predicted for 564 of those HPs. On the other hand, the MSR17 strain encoded 1228 HPs, and 514 of those encoded HPs indicated hits on the GO database. For the identification of protein domains and/or families, this pool of 1078 proteins was thoroughly investigated using the HMMER, Pfam, Superfamily,

Gene3D, and InterPro, NCBI Batch CD-Search tools. The results obtained from using those tools were analyzed to determine the appropriate functions to refer to HPs. Proteins with similar function predictions from three or even more programs were functionally annotated with high confidence. Consequently, the functions of 656 HPs (338 HPs from MSR16 strain and 318 HPs from MSR17 strain) have been annotated with a high level of confidence (Supplementary Table 2). The NCBI BLASTp program was used to validate the annotated functions of these HPs based on homologous proteins, indicating the accuracy of the annotation.

Analysis of GO terms of annotated HPs

In this study, the GO terms of 656 HPs with functional annotation were evaluated to identify their association with any of the following GO categories: Biological Process (BP), Cellular Components (CC), and Molecular Functions (MF). As far as the biological process is considered, there were 224 proteins identified with 78 GO keywords in total (Supplementary Table 3). This analysis revealed a cluster of transport proteins as well as the protein clusters essential for growth (Figure 2A). A total of 187 proteins were classified as cellular components using 18 different GO keywords. Surprisingly, 89 proteins were found to be integral component of the membrane (Figure 2B, Supplementary Table 4). In terms of molecular function, 370 proteins with 134 distinct GO terms were identified, and after analysis, a cluster of metal ion binding proteins was observed (Figure 2C, Supplementary Table 5). These findings imply that HPs might have a role in the development and pathogenesis of the organism, and the identified groups were investigated further.

Enzymes related with growth and survival. Within the set of annotated HPs, multiple enzymes that are essential to the growth and development of bacteria were observed. To have a complete grasp of the host-pathogen interaction, it is necessary to have knowledge about these enzymes. MSR16_hyp_1391 and MSR14_hyp_815 were annotated to be a radical SAM (S-adenosylmethionine) protein, whereas MSR16_hyp_11, MSR16_hyp_148, MSR16_hyp_1211, MSR17_hyp_124, MSR17_hyp_539, MSR17_hyp_768 and MSR17_hyp_797 were found to be SAM-dependent methyltransferase. It is well established that radical SAM proteins play a critical role in organism survival, and it has also been demonstrated that inhibiting these enzymes has been effective in preventing serious infections.[67,68] The phasin protein (MSR16_hyp_676 and MSR17_hyp_1221) was identified in the genome of the both strain. Phasins are the major polyhydroxyalkanoate (PHA) granule-associated proteins. They have both structural as well as regulatory functions, and they have the ability to affect the accumulation of PHA within the bacterial cell and to mediate protein folding, which in turn stimulates the growth of bacteria.[69,70]
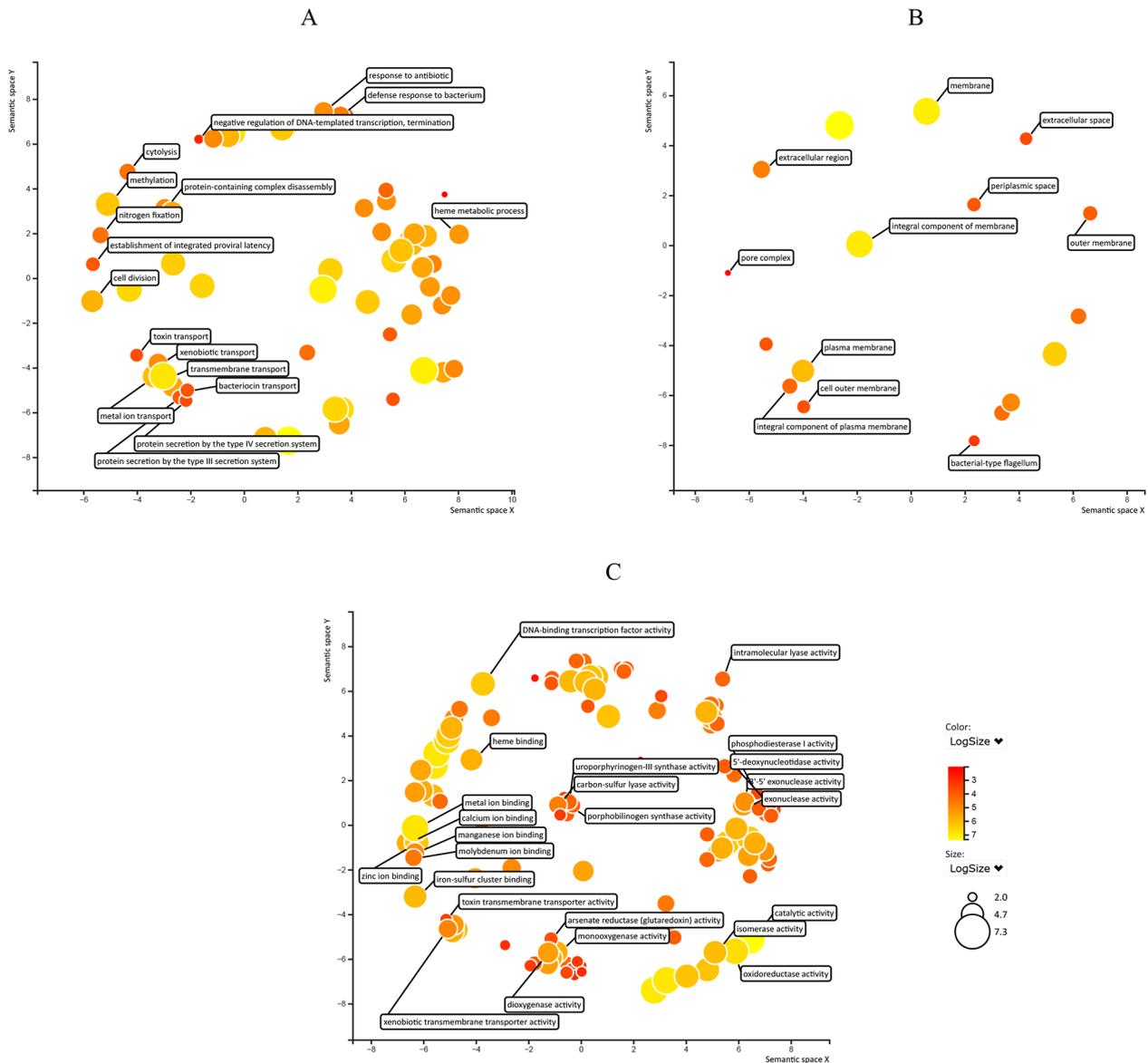
A

B

C

Figure 2. Classification of the HPs based on gene ontology data: (A) enriched biological processes, (B) enriched cellular components, and (C) enriched molecular function. The logSize of the enzyme number is represented both by the diameters of the circles and the colors used. HP indicates hypothetical protein.

The ubiquinone biosynthesis genes, which were encoded by both of the bacterial strains (MSR16_hyp_1380, MSR16_hyp_1387, MSR17_hyp_151, MSR17_hyp_161 and several others), were considered to be an additional significant cluster of genes that have been identified as a part of this study. Ubiquinone, commonly known as Coenzyme Q (CoQ), is a component of the respiratory chain in many prokaryotic organisms and is crucial for energy production as well as a variety of other intracellular activities.[71] Surprisingly, ubiquinone was found to be linked to bacterial pathogenicity in *Francisella novicida* and *Xanthomonas campestris*.[72,73]

Gram-negative bacteria have such a cell envelope that is made up of two membranes known as the inner membrane and the outer membrane (OM), along with an enclosed chamber that is known as the periplasm. The bacterial cell envelope is of particular interest because it serves the dual purpose as both a structural component as well as a permeability barrier.[74-76] These unique dual characteristics make the cell envelop a target of exhaustive study. Several HPs were identified in this study that may have a function in cell envelop biogenesis. OmpA (MSR16_hyp_1006 and MSR17_hyp_41) is a well-studied protein which is a major virulence factor that mediates the formation of bacterial biofilm, infection of eukaryotic cells, antibiotic resistance, and immunomodulation.[77]

For the synthesis of peptidoglycan in bacterial cell walls, the glycosyltransferase (MSR16_hyp_1127) is essential. This enzyme transfers the disaccharide-peptide from the lipid II onto the expanding glycan chain.[78] The peptidoglycan binding protein (MSR16_hyp_157, MSR17_hyp_1035, and MSR17_hyp_864) may also play a part in cell wall biogenesis.[79]

MSR16_hyp_229 and MSR17_hyp_548 were identified as BamA, which play a major role in the biogenesis of outer membrane (OM) proteins in bacteria.[80] An additional protein (MSR16_hyp_875), characterized as TolA, is engaged in the process of ensuring the integrity of the outer membrane.[81] To survive, it is extremely crucial for bacteria to monitor and preserve the integrity of the cell envelope in the presence of agents and situations that can destabilize the envelope.[82,83]

Interestingly, we found several potential therapeutic target proteins within the annotated HPs. Shikimate kinase (MSR16_hyp_748) is a possible therapeutic target that can be used against both of the strains that are being explored in this study. Studies are currently being conducted with the goal of identifying shikimate kinase inhibitors that are effective against methicillin-resistant *Staphylococcus aureus* and *Mycobacterium tuberculosis*.[84-86] The phosphoribosyl-ATP pyrophosphohydrolase (MSR16_hyp_1190), an enzyme that is involved in the histidine biosynthesis pathway, is another possible therapeutic target.[87] It has been experimentally validated that extracts from tropical plants have the ability to block aminoacyl-tRNA hydrolase (MSR16_hyp_1287 and MSR17_hyp_580), an enzyme that is essential to the process of protein biosynthesis.[88,89] Another possible target for novel selective antimicrobial drugs is the enzyme GTP cyclohydrolase-2 (MSR16_hyp_1336 and MSR17_hyp_1153). GTP cyclohydrolase-2 is involved in the process of flavin biosynthesis, and an inhibition of the flavin biosynthesis pathway may result in the death of the bacteria.[90]

When it comes to function or stability of a protein, the incorporation of disulfide bonds into proteins can be absolutely necessary. The periplasmic enzyme DsbA is essential for the incorporation of disulfide bonds into a large number of extracytoplasmic proteins and is associated with bacterial virulence.[91-94] We predicted that MSR16 hyp 605 is the DsbA oxidoreductase, and it has the potential to be a good therapeutic target.[95] The biosynthesis pathway for thiamine (vitamin B1) is yet another significant metabolic pathway in bacteria that possesses the potential to be exploited as a drug target.[96] The thiamine-phosphate synthase enzyme was observed to be present in both of the bacterial strains studied (MSR16_hyp_1301, MSR17_hyp_22, and MSR17_hyp_966). This enzyme can be targeted for drug development against the pathogens.[96,97]

Iron binding protein. During the process of pathogenesis, the ability of bacteria to accumulate iron and other essential metal ions from the surroundings becomes a major determining factor during the severity of the infection. This is due to the fact that hosts inevitably employ a strategy of nutritional immunity for minimizing the accessibility of metal ions to the bacteria.[98] Gram-negative bacteria comprise a variety of iron uptake systems that are capable of capturing iron-containing substrates to circumvent this situation.[99,100] One of the most interesting host-pathogen interactions to be examined is the competition for iron that takes place between pathogenic bacteria and the organisms that they infect.

Our pathogenic bacterial strains also encoded multiple proteins related to iron uptake and regulation, as expected. We found that bacterial strains in this study encoded the enzyme known as siderophore ferric iron reductase (MSR16_hyp_1279 and MSR17_hyp_934). Siderophores have a high affinity for ferric ions (Fe^{3+}), whereas they have only a moderate affinity for ferrous ions (Fe^{2+}).[101] Ferric iron reductase is an enzyme that catalyzes the reduction of ferric iron to ferrous, which is a form of iron that has a lower affinity and is highly soluble.[102,103] In addition, we observed that the pathogens encoded proteins essential for heme biosynthesis (MSR16_hyp_1339, MSR16_hyp_1340, MSR17_hyp_328 and MSR17_hyp_1130). To maintain the stability of the iron homeostasis within the bacterial cell during an infection, the synthesis of heme is a vital step.[104]

Antibiotic resistance genes. We identified several proteins which are associated with antibiotic resistance. A large cluster of MFS transporters was identified within the genome of both strains (for example- MSR16_hyp_1321, MSR16_hyp_1322, MSR16_hyp_1324, MSR17_hyp_162, MSR17_hyp_326, MSR17_hyp_475, MSR17_hyp_734, and few more). MFS transporters are capable of transporting a wide range of substrates across the cell wall, leading to the development of multidrug resistance (MDR) strains.[105,106] PACE efflux transporter (MSR16_hyp_1198) is a distinct type of transporter protein that is encoded by pathogens. Like other types of transporter proteins, it transports antimicrobial compounds out of the cell.[107]

GNAT family N-acetyltransferase proteins were also identified (MSR16_hyp_832, MSR16_hyp_878, MSR16_hyp_1039, MSR16_hyp_1067, MSR16_hyp_1109, MSR16_hyp_1326, and MSR16_hyp_1327 were encoded by MSR16 strain while MSR17_hyp_139, MSR17_hyp_200, MSR17_hyp_526, MSR17_hyp_573, and MSR17_hyp_1100 were encoded by MSR17 strain). GNAT family N-acetyltransferase proteins have prominent roles across a wide variety of biological processes, one of which is the development of aminoglycoside antibiotic resistance.[108,109]

We further identified the antibiotic resistance protein VanZ (MSR16_hyp_778 and MSR17_hyp_703), which reduces the binding of lipoglycopeptide antibiotics to cell wall components, resulting in resistance to teicoplanin and vancomycin antibiotics.[110,111]

Toxin proteins and toxin transporters. The thermostable direct hemolysin (TDH) and the TDH related hemolysin (TRH) are both regarded to be key virulence factors in pathogenic strains of the bacteria *V. parahaemolyticus*.[112,113] Along with them, we identified an additional hemolysin protein called enterohemolysin EhxA (MSR16_hyp_1044 and MSR17_hyp_988) that is

encoded in the genome of the bacteria and contributes to the pathogenesis of the bacteria.[114-116] In addition to this, we found an RTX toxin (MSR16_hyp_560) that is capable of acting as a synergistic virulence factor.[117-119] We also identified Hemolysin D (HlyD) (MSR16_hyp_510) and the outer membrane protein TolC (MSR16_hyp_929), which are involved in the hemolysin secretion system. The HlyD protein forms a continuous channel by docking to the protein TolC which forms a part of HlyA specific type I secretion system (T1SS).[120-122] We also identified two types of phospholipase; zinc-dependent phospholipase C (MSR16_hyp_651 and MSR16_hyp_988) and patatin-like phospholipase (MSR16_hyp_706, MSR17_hyp_54, and MSR17_hyp_854). Phospholipase C, commonly known as alpha-toxin, can bind to eukaryotic cell membranes and hydrolyze membrane lipid moieties (ie, phosphatidylcholine and sphingomyelin), leading to cell lysis.[123,124] Again, genomes of bacterial pathogens, particularly gram-negative species, encodes patatin-like PLA2 enzymes which act as effector molecules to target host cellular membranes, suggesting a role in host-pathogen interaction.[125,126] We also identified two other virulence genes in this study: VcgC (MSR16_hyp_127 and MSR17_hyp_103) and transcriptional activator HlyU (MSR16_hyp_114), which are commonly found in other *Vibrio* species also.[127,128]

Secretion system and associated proteins. It is widely known that gram-negative bacteria possess nine distinct types of secretion systems (type I-IX).[95,129] Out of them, the first six types (type I-VI) are the most prevalent and are linked to the virulence of the bacteria.[130] In our study, we also found several effector proteins as well as proteins involved in formation of secretion system.

We identified a large number of protein involved in the formation of type VI secretion system (eg, MSR16_hyp_1341, MSR16_hyp_1344, MSR16_hyp_1345, MSR17_hyp_1036, MSR17_hyp_1140, MSR17_hyp_1043, MSR17_hyp_1196 and several others). In gram-negative bacteria, Type VI secretion systems are associated with the machinery that is required for injecting effector proteins into eukaryotic cells to exert virulence, symbiosis, as well as antibacterial activity.[130-132] We also identified S-type Pyocin (MSR16_hyp_89 and MSR17_hyp_256), an effector of the type VI secretion system that has the potential to mediate antibacterial toxicity.[133,134]

V. parahaemolyticus also encodes genes for the formation of the type III secretion system, and we identified Type III secretion chaperone CesT among the hypothetical proteins (MSR16_hyp_195, MSR16_hyp_1271, MSR17_hyp_116, and MSR17_hyp_745). CesT chaperone is a multi-effector chaperone and is required for secretion of several effector proteins.[135-137] Another annotated HP which might be involved in bacterial virulence is the type II secretion system (T2SS) pilot lipoprotein GspSβ (MSR16_hyp_82 and MSR17_hyp_414). This protein is essential for the formation of the T2SS, thus facilitating the pathogenesis of the bacteria.[138,139]

Physicochemical parameters and subcellular localization of the annotated HPs

In this study, the amino acid sequences of all 657 annotated HPs were examined to determine their physicochemical properties (Supplementary Table 6). It is well recognized that in silico analysis could reveal molecular relevance since several parameters are related to protein stability and function. According to the ProtParam results, the length of the analyzed sequences ranged from 53 to 1578 amino acids. The theoretical pI of the proteins ranged from 3.68 to 10.55. This parameter refers to the point at which the amino acid can no longer tolerate liquid charge and the mobility of the ampholyte sums to zero.[140] The direction of protein migration on the gel during electrophoresis is determined by the charge. As a result, proteins can be separated in a gel based on their pI.[141] The molecular weights ranged from 5910.23 to 176889.35 Da. In laboratorial experiments, 2D gel electrophoresis visualization can be accompanied by a combination of pI and molecular weight, allowing new proteins to be identified and their relative abundance measured between comparative samples.[142]

For the HPs, the extinction coefficient, which is crucial for determining the accurate concentration of a protein, was also predicted. With respect to the concentration of cystine, tryptophan, and tyrosine amino acid residues in the protein sequences, the extinction coefficients of the proteins ranged from 125 M^{-1}cm^{-1} to 235185 M^{-1}cm^{-1} at 280 nm. High extinction coefficient occurred in some HPs because of the presence of high concentration of the cystine, tryptophan, and tyrosine residues while extinction coefficient was not predicted for some HPs because of the absence of these residue in their protein sequence.

The aliphatic index and instability index were predicted to evaluate the stability of HPs. The aliphatic index is directly related to the molecular fraction of aliphatic amino acids (ie, alanine, valine, isoleucine, and leucine) and is related to the thermostability of the protein; the higher the aliphatic index, the greater the thermostability.[143] Values for aliphatic index ranged from 44.86 to 155.24. The instability index was used to estimate an assumption of protein stability in a test tube. Proteins with instability index less than 40 are regarded as stable proteins.[144] In the study, instability index for the annotated HPs ranged from 10.19 to 98.97, indicating that 367 out of 656 proteins should be stable in a test tube.

Finally, the GRAVY (Grand Average of Hydropathy) value of the proteins was predicted. This value ranged from –0.535 to 0.434 in the study, with 543 proteins having a score of less than 0 and 113 proteins having a score of greater than 0. Proteins that have GRAVY values that are lower than 0 are regarded as relatively hydrophilic, whereas proteins that have GRAVY scores that are higher than 0 are regarded as relatively hydrophobic.[145] This information is useful for identifying proteins by categorizing them as globular or membrane-bound proteins.[145,146]

Since the function of a protein is typically associated with the location of the protein, the subcellular localization of a
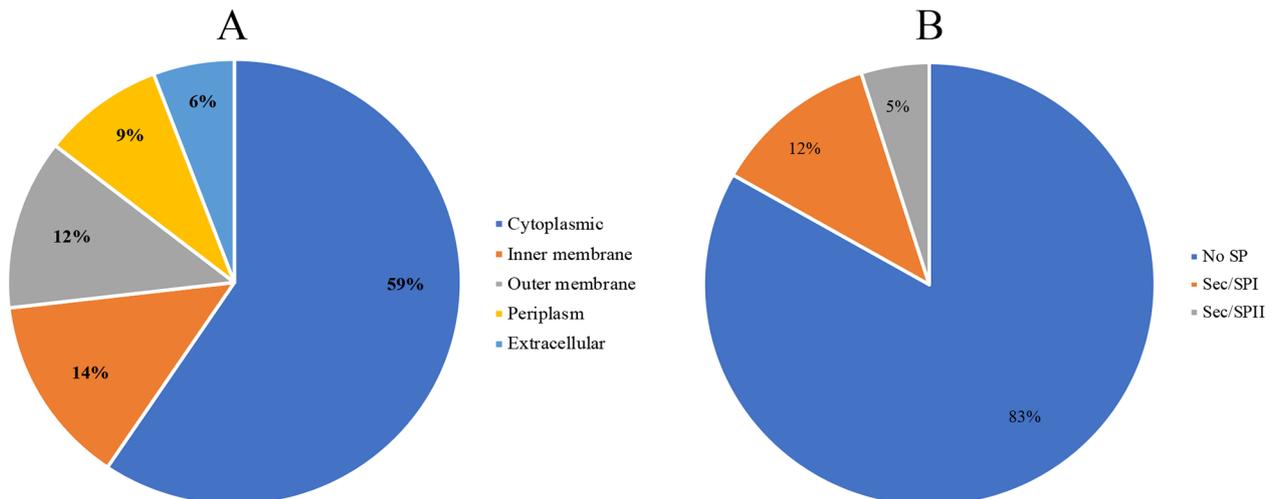
A

B

Figure 3. Annotated HPs classified on the basis of (A) subcellular localization and (B) signal peptide. HP indicates hypothetical protein.

protein can provide helpful insights regarding the functions of proteins.[147,148] Among the annotated HPs, 60% (390) of them were predicted to be localized at the cytoplasm. The number of HPs present at the inner membrane 17.41% (90) and 2.52% (81) are present at outer membrane. It is predicted that 14% (57) HPs are present in periplasm and 5.89% (38) in extra cellular matrix (Figure 3A, Supplementary Table 7).

Signal peptides are the key players in determining the transport of proteins to the target location. Hence, prediction of signal peptide is essential to learn about the transport system of the specific proteins and the cleavage sites. We predicted the presence of signal peptide sequences in 110 HPs out of 656 (78 contains Sec/SPI and 32 HPs contain Sec/SPII) (Figure 3B, Supplementary Table 7). Presence of transmembrane helices were also analyzed using the protein sequences, and it was predicted that transmembrane helices were present among 120 HPs (Supplementary Table 7).

Identification of virulent proteins

MP3, VirulentPred, and BastionHub were employed to accurately predict virulence factors with a high level of confidence. Only those proteins that had been projected to be pathogenic by each of the three different prediction tools were selected and studied further. There were a total of 656 annotated HPs, of which 92 were predicted to be virulent. This included 47 HPs from the MSR16 strain and 45 HPs from the MSR17 strain (Supplementary Table 8).

Virulent HPs with therapeutic potential. The antigenicity of the pathogenic HPs was analyzed, and the results indicated that 56 of them have the ability to produce an antigenic response, and out of those 56, 36 proteins have been projected to be non-allergenic and non-toxic toxic (Supplementary Table 8). These results indicate that these proteins could be used as vaccine candidates. Further analysis of these proteins identified that 27 of the candidate proteins either are extracellular proteins or are

localized at the outer membrane or periplasm of the bacterium. This implies that these proteins may have the potential to be used in the development of a vaccine that is effective against the pathogen (Table 1).

PPI network analysis

On the annotated HPs, we performed a PPI analysis, which is a method that can further validate the annotation and also suggest potential roles for these proteins. We had to exclude 161 HP from a total of 656 due to low confidence interactions with other proteins. In terms of the remaining 495 HPs, 65 of them either have domains with an unidentified function or the protein itself has not been defined. As a result, 430 proteins have been identified and clustered according to their functions (Supplementary Table 9). These functions include DNA mismatch repair enzymes, DNA/RNA binding molecules, transcription factors, multidrug resistance protein, nitrogen fixation protein, and several other. These proteins have the potential to play a major role in the growth and survival of the pathogen. Target HPs are indicated by homologous proteins in *V. parahaemolyticus* RIMD 2210633 strain (Supplementary Table 9).

We performed a comprehensive analysis of the protein clusters to identify proteins that play a significant role in the development and pathogenesis of the pathogen. Within the scope of this study, we were able to reveal an interconnected clusters of proteins that play essential roles in the biosynthesis of small compounds (Figure 4A). These compounds include thiamine, heme, riboflavin, and folic acid. Three of these identified pathways, which are encoded by the pathogen of interest, do not exist in mammals. Since mammals do not have the enzymes required for thiamine biosynthesis, the enzymes that are involved in this process could be a viable target for the development of drugs to treat this infection.[149] In addition to thiamine, riboflavin is an essential biomolecule for the continued survival of an organism. However, like thiamine, mammals do not encode the enzymes necessary for the production of riboflavin,

Table 1. List of virulent HPs with therapeutic potential.

VIRULENT HP	ANTIGENICITY	ALLERGENICITY	TOXICITY	SUBCELLULAR LOCALIZATION
MSR16_hyp_1044	Antigen	Non-allergen	Non-toxin	Extracellular
MSR16_hyp_1061	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR16_hyp_1113	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR16_hyp_1143	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR16_hyp_1310	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR16_hyp_1339	Antigen	Non-allergen	Non-toxin	Periplasmic
MSR16_hyp_216	Antigen	Non-allergen	Non-toxin	Periplasmic
MSR16_hyp_229	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR16_hyp_327	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR16_hyp_560	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR16_hyp_626	Antigen	Non-allergen	Non-toxin	Extracellular
MSR16_hyp_631	Antigen	Non-allergen	Non-toxin	Extracellular
MSR16_hyp_698	Antigen	Non-allergen	Non-toxin	Extracellular
MSR16_hyp_874	Antigen	Non-allergen	Non-toxin	Extracellular
MSR17_hyp_1130	Antigen	Non-allergen	Non-toxin	Periplasmic
MSR17_hyp_186	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR17_hyp_319	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR17_hyp_345	Antigen	Non-allergen	Non-toxin	Periplasmic
MSR17_hyp_392	Antigen	Non-allergen	Non-toxin	Periplasmic
MSR17_hyp_406	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR17_hyp_512	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR17_hyp_545	Antigen	Non-allergen	Non-toxin	Extracellular
MSR17_hyp_548	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR17_hyp_744	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR17_hyp_790	Antigen	Non-allergen	Non-toxin	Extracellular
MSR17_hyp_831	Antigen	Non-allergen	Non-toxin	Outer membrane
MSR17_hyp_988	Antigen	Non-allergen	Non-toxin	Extracellular

Abbreviation: HP, hypothetical protein.

which demonstrates their suitability for use in the discovery of new drugs.[150] The potential druggability of the folic acid biosynthetic pathway for the development of therapeutic intervention has been experimentally validated by the use of drugs inhibiting folic acid biosynthesis.[151,152]

A protein cluster that can bind to DNA was also observed (Figure 4B). Proteins necessary for DNA synthesis, DNA mismatch repair, and homologous recombination have been found in this cluster. Virulence proteins play a pivotal role in pathogenesis. There are two ways that virulence proteins can be transported: they can either be bound to the membrane or

secreted through the secretion system. In this study, we were able to identify proteins that are necessary for the production of biofilms, in addition to components of the type VI secretion system (T6SS) (Figure 4C).

Discussion

A significant portion of the proteomes of both prokaryotic and eukaryotic organisms are made up of hypothetical proteins (HP).[153] Hypothetical proteins are proteins that are assumed to be expressed from an open reading frame (ORF), although they have no experimental evidence of translation. Given the
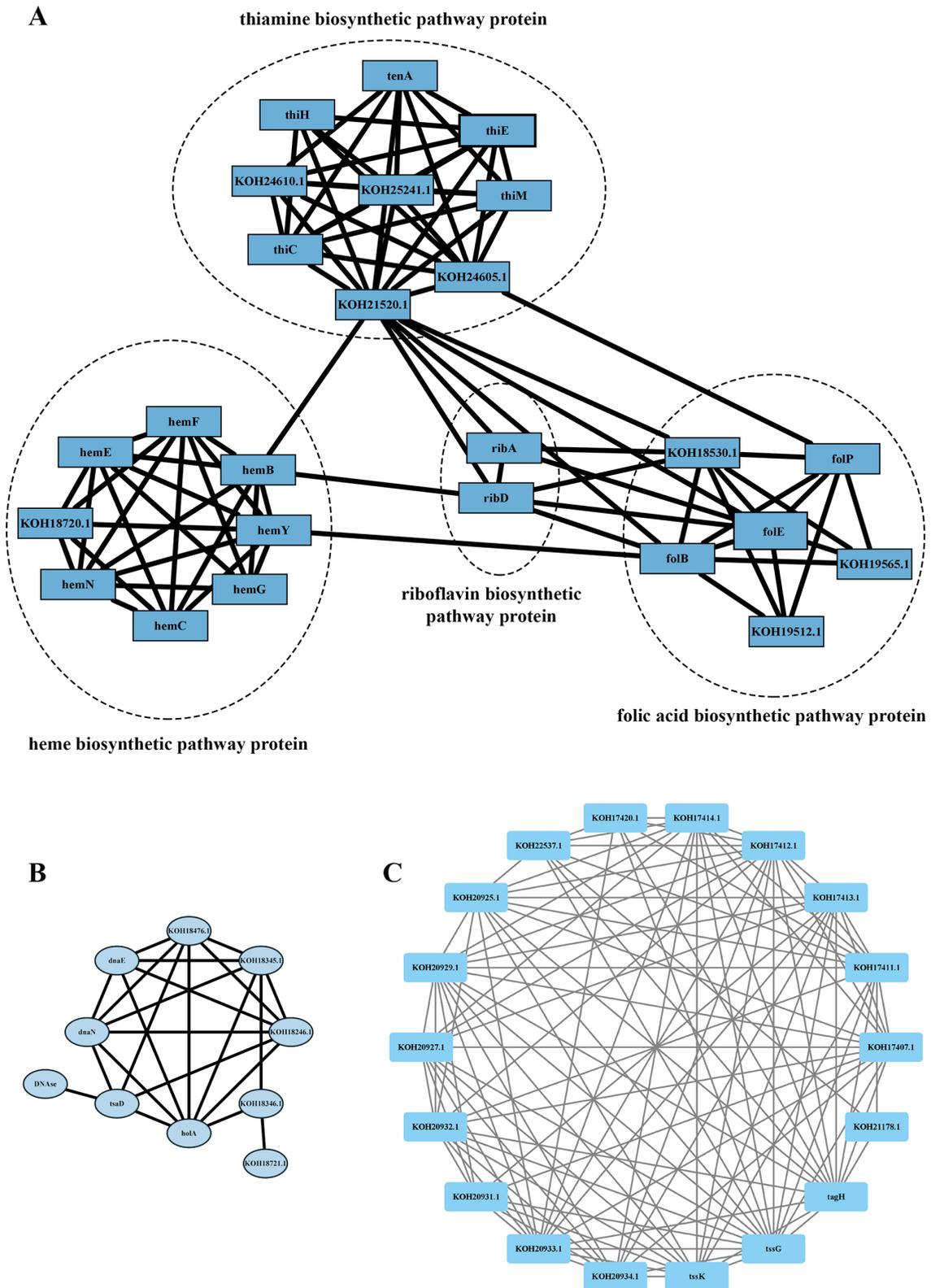
Figure 4. Protein-protein interaction analysis of the annotated HPs. The identified protein clusters include proteins involved in (A) biosynthesis of small compound, (B) DNA binding, and (C) biofilm formation. HP indicates hypothetical protein.

significance of understanding the underlying molecular mechanisms that are present in a variety of species, notably pathogenic bacteria, a number of studies have emphasized in past few years on the annotation of proteins that have not yet been assigned a function.[154,155] The wet lab approaches that are commonly employed for unraveling desired genes and proteins are time-consuming and expensive. As a result, the in silico methods have arisen as crucial tools for identification of the

hypothetical proteins and assigning their function.[21,156-158] Wet lab experiments have been used to validate the reliability of the in silico functional annotation approach.[154] An integrated in silico and in vivo method was used to functionally annotate the hypothetical proteins, which led to the identification of the proteins of *Pseudomonas* sp. Lz4W involved in cold adaptation[158] and also the high arsenic resistance genes from *Exiguobacterium antarcticum* strain B7.[21] These evidences have provided the impetus for the development of computational tools and techniques that have a higher degree of accuracy in elucidating the functions of HPs.

The purpose of this study was to annotate the HPs with the aim of filling the information gaps of the genome sequence analysis through the use of computational approaches. Identification of protein family and domain, subcellular localization prediction, secretome analysis, and protein-protein interaction analysis were among the approaches used. The development of new computational tools and the increased availability of existing ones have both had a favorable impact on the understanding of the correlation between the structure and function of proteins. In addition, these tools provided a basis of in vitro assays for the characterization of proteins. The urgency of properly understanding the roles that the hypothetical proteins play in the pathogenesis of *V. parahaemolyticus* and refining the functional annotation for future research makes this study essential.

To improve the accuracy of computational predictions, we devised a functional annotation pipeline based on sequence analysis coupled with the implementation of other approaches such as PPI (Figure 1). The conservation of structural properties may imply a function for some HP, which can then be annotated according to that function. These structural properties were identified through the integration of data that was available from multiple databases, evaluating conserved domains, families, and superfamilies. As a conclusion, using the pipeline, the functions of 656 of the 2631 HP have been annotated (Supplementary Table 2). We found several proteins that could be exploited as therapeutic and vaccine targets among the annotated HPs.

We discovered a varied array of proteins in the HPs by analyzing their GO keywords, ranging from enzymes essential for survival to structural components of the secretion system critical for virulence (Figure 2). These proteins are essential for a variety of cellular activities, and they also have the ability to play a role in the growth and survival of the organism. Enzymes are an essential component for the continued survival of an organism. In the case of bacteria, enzymes are not only important for promoting the growth and development of bacteria within their host by providing the necessary nutrients, but they are also responsible for the pathogenesis of infections. Enzymes have the ability to alter the local environment in such a way that it becomes favorable for the growth of bacteria as well as the metabolism of compounds that are available within the host.[159]

Thiamin, also known as vitamin B1, is an essential cofactor required for the metabolism of carbohydrates as well as branched-chain amino acids when it is present in its active form, thiamin diphosphate.[160,161] Although the majority of bacteria, as well as fungi and plants, are capable of de novo synthesis of thiamin, animals are incapable of thiamine synthesis and therefore must obtain essential thiamin entirely through their diet.[162] Recent research has pointed to the thiamin biosynthesis pathway of *Plasmodium falciparum*, the etiological agent causing tropical malaria, as a potential source of therapeutic targets.[163,164] It stands to reason that the enzymes involved in the biosynthesis of thiamine and the regulatory network that underlies this process should be good candidates for use as therapeutic targets in the development of novel antibacterial agents. We were able to identify the enzyme known as thiamin phosphate synthase (ThiE), a potential drug target which is responsible for catalyzing the coupling of thiazole with pyrimidine.[165] In a positive context, we identified another enzyme, GTP cyclohydrolase-2, that is essential for the survival of bacteria. This enzyme is involved in the pathway leading to the biosynthesis of riboflavin (Vitamin B2) and is a prominent antimicrobial drug target since animals do not possess this enzyme.[90,166]

During the characterization of these two strains, it was surprising to find that both of them were resistant to a wide range of antibiotics.[167,168] Both of the strains that were identified were able to demonstrate resistance to a wide variety of antibiotics, which may be explained by the presence of a large cluster of MFS proteins as well as GNAT family N-acetyltransferase proteins within their genome. This observation may shed light on how the strains gained this resistance. It can be hypothesized that the presence of a large number of multidrug resistance proteins and their expression have facilitated the evaluation of the pathogen to adopt mechanisms to demonstrate broad spectrum antibiotic resistance.[169,170]

In addition to the TDH and TRH, there were a few other hemolysin toxin proteins found in this study. This is an important observation that warrants more investigation. Enterohemolysin EhxA and phospholipase C were found to be present, and both of these proteins are essential for evading the tissue system of the host.[115,171-174]

Since virulence factors facilitate the colonization of the pathogen at the cellular level of the host and cause disease by evading the host defense mechanisms, it is necessary to have an understanding of the biological function and mechanism of virulence factors to comprehend the role they play during the pathogenesis of bacteria.[175,176] Also, in the case of bacterial infections, virulent factors (Supplementary Table 8) have the potential to be used as drug target.[177,178] When combined with antibiotics, annotated virulent HPs can contribute to the development of a more effective target-based drug discovery approach and facilitate the treatment of bacterial infections. In the event that these virulence factors are able to circumvent the

host's immune system, the final outcome will be either the sufficient multiplication to establish an infection or persistence of the microbe within the host tissue. It is possible that this will either cause considerable damage to the tissue of the host or enable transmission of the pathogen to other susceptible hosts.[179] Small molecule inhibitors and antibodies are two examples of the types of strategies that can be employed to suppress the activity of the virulence factors.[180] Consequently, the development of a vaccine that targets the virulence factors of these pathogens might be an additional strategy for tackling them.[180,181] Throughout the course of our research, we were able to identify a number of virulence factors that hold promise as future areas of concentration for vaccine research (Table 1). The characterization of these proteins warrants for additional research to be conducted.

In the current study, functional and physicochemical characteristics of the HP that are encoded by two pathogenic strains of *V. parahaemolyticus* were able to be elucidated. This allowed for a better understanding of the role that the HPs play during the growth and pathogenesis of the bacteria. It is conceivable, in context of the molecular mechanisms underlying bacterial pathogenesis, that the virulent proteins reported in this study induce hemolysis and cell damage, which result in AHPND and shrimp mortality.

The findings of this study have provided significant information about proteins that are present within the two strains of this pathogen but roles were previously unknown. Our findings lead to the possibility of further research involving these targets, particularly to explore the role of these proteins in vivo and/or in vitro, as well as their potential to play a role in pathogenesis and their application for the development of novel therapeutic targets against *V. parahaemolyticus*.

Acknowledgements

Author Contributions

ABMMKI conceived the project. SS collected the data. SS, MAZ, and ABMMKI performed the analyses. SS, MS, and ABMMKI wrote the manuscript. The manuscript was reviewed and approved by all authors.

Data and Software Availability

All data are added in the table, figures, and supplementary file and supplementary tables. In this research work, publicly available, free, mostly online and few offline software/tools were used. Necessary link and reference of the software/tools are provided in the method section.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Kumar R, Ng TH, Wang H. Acute hepatopancreatic necrosis disease in penaeid shrimp. *Rev Aquac*. 2020;12:1867-1880. doi:10.1111/raq.12414.

2. Lightne DV. The penaeid shrimp viruses TSV, IHHNV, WSSV, and YHV. *J Appl Aquac*. 1999;9:27-52. doi:10.1300/J028v09n02_03.

3. Kusumaningrum HP, Zainuri M. Detection of bacteria and fungi associated with *Penaeus monodon* postlarvae mortality. *Procedia Environ Sci*. 2015;23:329-337. doi:10.1016/j.proenv.2015.01.048.

4. Ina-Salwany MY, Al-Saari N, Mohamad A, et al. Vibriosis in fish: a review on disease development and prevention. *J Aquat Anim Health*. 2019;31:3-22. doi:10.1002/aah.10045.

5. Hossain MMM, Uddin MI, Islam H, et al. Diagnosis, genetic variations, virulence, and toxicity of AHPND-positive *Vibrio parahaemolyticus* in *Penaeus monodon*. *Aquac Int*. 2020;28:2531-2546. doi:10.1007/s10499-020-00607-z.

6. Tran L, Nunan L, Redman R, et al. Determination of the infectious nature of the agent of acute hepatopancreatic necrosis syndrome affecting penaeid shrimp. *Dis Aquat Organ*. 2013;105:45-55. doi:10.3354/dao02621.

7. Shinn AP. Asian shrimp production and the economic costs of disease. *Asian Fish Sci*. 2018;31S:29-58. doi:10.33997/j.afs.2018.31.S1.003.

8. De Schryver P, Defoirdt T, Sorgeloos P. Early mortality syndrome outbreaks: a microbial management issue in shrimp farming? *PLoS Pathog*. 2014;10:e1003919. doi:10.1371/journal.ppat.1003919.

9. Lai H-C, Ng TH, Ando M, et al. Pathogenesis of acute hepatopancreatic necrosis disease (AHPND) in shrimp. *Fish Shellfish Immunol*. 2015;47:1006-1014. doi:10.1016/j.fsi.2015.11.008.

10. Sirikharin R, Taengchaiyaphum S, Sanguanrut P, et al. Characterization and PCR detection of binary, pir-like toxins from *Vibrio parahaemolyticus* isolates that cause acute hepatopancreatic necrosis disease (AHPND) in shrimp. *PLoS ONE*. 2015;10:e0126987. doi:10.1371/journal.pone.0126987.

11. Silvester R, Alexander D, Ammanamveetil MHA. Prevalence, antibiotic resistance, virulence and plasmid profiles of *Vibrio parahaemolyticus* from a tropical estuary and adjoining traditional prawn farm along the southwest coast of India. *Ann Microbiol*. 2015;65:2141-2149. doi:10.1007/s13213-015-1053-x.

12. Siddique AB, Moniruzzaman M, Ali S, et al. Characterization of pathogenic *Vibrio parahaemolyticus* isolated from fish aquaculture of the southwest coastal area of Bangladesh. *Front Microbiol*. 2021;12. doi:10.3389/fmicb.2021.635539.

13. Yeung PSM, Boor KJ. Epidemiology, pathogenesis, and prevention of foodborne *Vibrio parahaemolyticus* infections. *Foodborne Pathog Dis*. 2004;1:74-88. doi:10.1089/153531404323143594.

14. Wu Y, Wen J, Ma Y, Ma X, Chen Y. Epidemiology of foodborne disease outbreaks caused by *Vibrio parahaemolyticus*, China, 2003-2008. *Food Contr*. 2014;46:197-202. doi:10.1016/j.foodcont.2014.05.023.

15. Han J, Tang K, Tran L, Lightner D. Photorhabdus insect-related (Pir) toxin-like genes in a plasmid of *Vibrio parahaemolyticus*, the causative agent of acute hepatopancreatic necrosis disease (AHPND) of shrimp. *Dis Aquat Organ*. 2015;113:33-40. doi:10.3354/dao02830.

16. Lee C-T, Chen I-T, Yang Y-T, et al. The opportunistic marine pathogen *Vibrio parahaemolyticus* becomes virulent by acquiring a plasmid that expresses a deadly toxin. *Proc Natl Acad Sci*. 2015;112:10798-10803. doi:10.1073/pnas.1503129112.

17. Li P, Kinch LN, Ray A, et al. Acute hepatopancreatic necrosis disease-causing *Vibrio parahaemolyticus* strains maintain an antibacterial type VI secretion system with versatile effector repertoires. *Appl Environ Microbiol*. 2017;83. doi:10.1128/AEM00737-17.

18. Ahmmed S, Khan MA-A-K, Eshik MME, Punom NJ, Islam ABMMK, Rahman MS. Genomic and evolutionary features of two AHPND positive *Vibrio parahaemolyticus* strains isolated from shrimp (*Penaeus monodon*) of south-west Bangladesh. *BMC Microbiol*. 2019;19:270. doi:10.1186/s12866-019-1655-8.

19. Luangtrakul W, Boonchuen P, Jaree P, Kumar R, Wang HC, Somboonwiwat K. Cytotoxicity of *Vibrio parahaemolyticus* AHPND toxin on shrimp hemocytes, a newly identified target tissue, involves binding of toxin to aminopeptidase N1 receptor. *PLoS Pathog*. 2021;17:e1009463. doi:10.1371/journal.ppat.1009463.

20. Pui CF, Bilung LM, Bainun N, et al. Risk of acquiring *Vibrio parahaemolyticus* in water and shrimp from an aquaculture farm. *Kuroshio Sci*. 2014;8:59-62. https://pdfs.semanticscholar.org/e853/d4294ab729993cf57c1c684bc4b93fe14abc.pdf.

21. da Costa WLO, de A Araújo CL, Dias LM, et al. Functional annotation of hypothetical proteins from the *Exiguobacterium antarcticum* strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. *PLoS ONE*. 2018;13:e0198965. doi:10.1371/journal.pone.0198965.

22. Galperin MY. Conserved "hypothetical" proteins: new hints and new puzzles. *Comp Funct Genomics*. 2001;2:14-18. doi:10.1002/cfg.66.

23. Eisenstein E, Gilliland GL, Herzberg O, et al. Biological function made crystal clear—annotation of hypothetical proteins via structural genomics. *Curr Opin Biotechnol*. 2000;11:25-30. doi:10.1016/S0958-1669(99)00063-4.

24. Pranavathiyani G, Prava J, Rajeev AC, Pan A. Novel target exploration from hypothetical proteins of *Klebsiella pneumoniae* MGH 78578 reveals a protein

involved in host-pathogen interaction. *Front Cell Infect Microbiol*. 2020;10. doi:10.3389/fcimb.2020.00109.

25. Shahbaaz M, Hassan MI, Ahmad F. Functional annotation of conserved hypothetical proteins from *Haemophilus influenzae* Rd KW20. *PLoS ONE*. 2013;8:e84263. doi:10.1371/journal.pone.0084263.

26. Araújo CL, Blanco I, Souza L, et al. In silico functional prediction of hypothetical proteins from the core genome of *Corynebacterium pseudotuberculosis* biovar ovis. *PeerJ*. 2020;8:e9643. doi:10.7717/peerj.9643.

27. Rao VS, Srinivas K, Sujini GN, Kumar GN. Protein-protein interaction detection: methods and analysis. *Int J Proteomics*. 2014;2014:147648-147612. doi:10.1155/2014/147648.

28. Agarwala R, Barrett T, Beck J, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2018;46:D8-D13. doi:10.1093/nar/gkx1095.

29. Araujo FA, Barh D, Silva A, Guimarães L, Ramos RTJ. GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. *Sci Rep*. 2018;8:1794. doi:10.1038/s41598-018-20211-9.

30. Törönen PP, Holm L. PANNZER—a practical tool for protein function prediction. *Protein Sci*. 2022;31:118-128. doi:10.1002/pro.4193.

31. Götz S, García-Gómez JM, Terol J, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008;36:3420-3535. doi:10.1093/nar/gkn176.

32. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res*. 2018;46:W200-W204. doi:10.1093/nar/gky448.

33. Mistry J, Chuguransky S, Williams L, et al. Pfam: the protein families database in 2021. *Nucleic Acids Res*. 2021;49:D412-D419. doi:10.1093/nar/gkaa913.

34. Pandurangan AP, Stahlhacke J, Oates ME, Smithers B, Gough J. The SUPER-FAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res*. 2019;47:D490-D494. doi:10.1093/nar/gky1130.

35. Lewis TE, Sillitoe I, Dawson N, et al. Gene3D: extensive prediction of globular domains in proteins. *Nucleic Acids Res*. 2018;46:D435-D439. doi:10.1093/nar/gkx1069.

36. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30:1236-1240. doi:10.1093/bioinformatics/btu031.

37. Mitchell AL, Attwood TK, Babbitt PC, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*. 2019;47:D351-D360. doi:10.1093/nar/gky1100.

38. Lu S, Wang J, Chitsaz F, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. 2020;48:D265-D268. doi:10.1093/nar/gkz991.

39. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinform*. 2015;16:169. doi:10.1186/s12859-015-0611-3.

40. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008;36:W5-W9. doi:10.1093/nar/gkn201.

41. Boekhorst J, Snel B. Identification of homologs in insignificant blast hits by exploiting extrinsic gene properties. *BMC Bioinform*. 2007;8:356. doi:10.1186/1471-2105-8-356.

42. Wan X-F, Xu D. Computational methods for remote homolog identification. *Curr Protein Pept Sci*. 2005;66:527-546. doi:10.2174/138920305774933231.

43. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*. 2011;6:e21800. doi:10.1371/journal.pone.0021800.

44. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press;2005;571-607. doi:10.1385/1-59259-890-0:571.

45. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: a fresh approach to numerical computing. *SIAM Rev*. 2017;59:65-98. doi:10.1137/141000671.

46. Yu NY, Wagner JR, Laird MR, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*. 2010;26:1608-1615. doi:10.1093/bioinformatics/btq249.

47. Yu C-S, Lin C-J, Hwang JK. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci*. 2004;13:1402-1406. doi:10.1110/ps.03479604.

48. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305:567-580. doi:10.1006/jmbi.2000.4315.

49. Dobson L, Reményi I, Tusnády GE. CCTOP: a consensus constrained TOPology prediction web server. *Nucleic Acids Res*. 2015;43:W408-W412. doi:10.1093/nar/gkv451.

50. Teufel F, Almagro Armenteros JJ, Johansen AR, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol*. 2022;40:1023-1025. doi:10.1038/s41587-021-01156-3.

51. Gupta A, Kapil R, Dhakan DB, Sharma VK. MP3: a software tool for the prediction of pathogenic proteins in genomic and metagenomic data. *PLoS ONE*. 2014;9:e93907. doi:10.1371/journal.pone.0093907.

52. Garg A, Gupta D. VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. *BMC Bioinformatics*. 2008;9:62. doi:10.1186/1471-2105-9-62.

53. Wang J, Li J, Hou Y, et al. BastionHub: a universal platform for integrating and analyzing substrates secreted by Gram-negative bacteria. *Nucleic Acids Res*. 2021;49:D651-D659. doi:10.1093/nar/gkaa899.

54. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007;8:1-7. doi:10.1186/1471-2105-8-4.

55. Magnan CN, Zeller M, Kayala MA, et al. High-throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics*. 2010;26:2936-2943. doi:10.1093/bioinformatics/btq551.

56. Wei L, Ye X, Sakurai T, Mu Z, Wei L. ToxIBTL: prediction of peptide toxicity based on information bottleneck and transfer learning. *Bioinformatics*. 2022;38:1514-1524. doi:10.1093/bioinformatics/btac006.

57. Maurer-Stroh S, Krutz NL, Kern PS, et al. AllerCatPro—prediction of protein allergenicity potential from the protein sequence. *Bioinformatics*. 2019;35:3020-3027. doi:10.1093/bioinformatics/btz029.

58. Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2021;49:D605-D612. doi:10.1093/nar/gkaa1074.

59. Folador EL, de Carvalho PVSD, Silva WM, et al. In silico identification of essential proteins in *Corynebacterium pseudotuberculosis* based on protein-protein interaction networks. *BMC Syst Biol*. 2016;10:103. doi:10.1186/s12918-016-0346-4.

60. Ngounou Wetie AG, Sokolowska I, Woods AG, Roy U, Deinhardt K, Darie CC. Protein-protein interactions: switch from classical methods to proteomics and bioinformatics-based approaches. *Cell Mol Life Sci*. 2014;71:205-228. doi:10.1007/s00018-013-1333-1.

61. Sharan R, Suthram S, Kelley RM, et al. Conserved patterns of protein interaction in multiple species. *Proc Natl Acad Sci USA*. 2005;102:1974-1979. doi:10.1073/pnas.0409522102.

62. Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013;41:D808-D815. doi:10.1093/nar/gks1094.

63. Huerta-Cepas J, Szklarczyk D, Forslund K, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res*. 2016;44:D286-D293. doi:10.1093/nar/gkv1248.

64. Yu H, Luscombe NM, Lu HX, et al. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res*. 2004;14:1107-1118. doi:10.1101/gr.1774904.

65. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498-2504. doi:10.1101/gr.1239303.

66. Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T, Albrecht M. Computing topological parameters of biological networks. *Bioinformatics*. 2008;24:282-284. doi:10.1093/bioinformatics/btm554.

67. Sofia HJ. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res*. 2001;29:1097-1106. doi:10.1093/nar/29.5.1097.

68. Parveen N, Cornell KA. Methylthioadenosine/S-adenosylhomocysteine nucleosidase, a critical enzyme for bacterial metabolism. *Mol Microbiol*. 2011;79:7-20. doi:10.1111/j.1365-2958.2010.07455.x.

69. Mezzina MP, Pettinari MJ. Phasins, multifaceted polyhydroxyalkanoate granule-associated proteins. *Appl Environ Microbiol*. 2016;82:5060-5067. doi:10.1128/AEM01161-16.

70. Zhao H, Wei H, Liu X, et al. Structural insights on PHA binding protein PhaP from *Aeromonas hydrophila*. *Sci Rep*. 2016;6:39424. doi:10.1038/srep39424.

71. Pierrel F, Burgardt A, Lee J-H, Pelosi L, Wendisch VF. Recent advances in the metabolic pathways and microbial production of coenzyme Q. *World J Microbiol Biotechnol*. 2022;38:58. doi:10.1007/s11274-022-03242-3.

72. Kazemzadeh K, Hajj Chehade M, Hourdoir G, et al. The biosynthetic pathway of ubiquinone contributes to pathogenicity of *Francisella novicida*. *J Bacteriol*. 2021;203. doi:10.1128/JB00400-21.

73. Zhou L, Li M, Wang XY, et al. Biosynthesis of coenzyme Q in the phytopathogen *Xanthomonas campestris* via a yeast-like pathway. *Mol Plant Microbe Interact*. 2019;32:217-226. doi:10.1094/MPMI-07-18-0183-R.

74. Mandela E, Stubenrauch CJ, Ryoo D, et al. Adaptation of the periplasm to maintain spatial constraints essential for cell envelope 2 processes and cell viability. *eLife*. 2022;11:1-21. doi:10.7554/eLife.73516.

75. Wang X, Bernstein HD. The *Escherichia coli* outer membrane protein OmpA acquires secondary structure prior to its integration into the membrane. *J Biol Chem*. 2022;298:101802. doi:10.1016/j.jbc.2022.101802.

76. Zgurskaya HI, López CA, Gnanakaran S. Permeability barrier of gram-negative cell envelopes and approaches to bypass it. *ACS Infect Dis*. 2015;1:512-522. doi:10.1021/acsinfecdis.5b00097.

77. Nie D, Hu Y, Chen Z, et al. Outer membrane protein A (OmpA) as a potential therapeutic target for *Acinetobacter baumannii* infection. *J Biomed Sci*. 2020;27:26. doi:10.1186/s12929-020-0617-7.

78. Mesleh MF, Rajaratnam P, Conrad M, et al. Targeting bacterial cell wall peptidoglycan synthesis by inhibition of glycosyltransferase activity. *Chem Biol Drug Des*. 2016;87:190-199. doi:10.1111/cbdd.12662.

79. Dörr T, Lam H, Alvarez L, Cava F, Davis BM, Waldor MK. A novel peptidoglycan binding protein crucial for PBP1A-mediated cell wall biogenesis in *Vibrio cholerae*. *Plos Genet*. 2014;10:e1004433. doi:10.1371/journal.pgen.1004433.

80. Doyle MT, Bernstein HD. Bacterial outer membrane proteins assemble via asymmetric interactions with the BamA β-barrel. *Nat Commun*. 2019;10:3358. doi:10.1038/s41467-019-11230-9.

81. Lazzaroni JC, Germon P, Ray MC, Vianney A. The Tol proteins of *Escherichia coli* and their involvement in the uptake of biomolecules and outer membrane stability. *FEMS Microbiol Lett*. 1999;177:191-197. doi:10.1016/S0378-1097(99)00293-1.

82. Jordan S, Hutchings MI, Mascher T. Cell envelope stress response in gram-positive bacteria. *FEMS Microbiol Rev*. 2008;32:107-146. doi:10.1111/j.1574-6976.2007.00091.x.

83. Wu Y, Bai J, Zhong K, et al. Antibacterial activity and membrane-disruptive mechanism of 3-p-trans-coumaroyl-2-hydroxyquinic acid, a novel phenolic compound from pine needles of *Cedrus deodara*, against *Staphylococcus aureus*. *Molecules*. 2016;21. doi:10.3390/molecules21081084.

84. Nunes JES, Duque MA, de Freitas TF, et al. Mycobacterium tuberculosis shikimate pathway enzymes as targets for the rational design of anti-tuberculosis drugs. *Molecules*. 2020;25:1259. doi:10.3390/molecules25061259.

85. Simithy J, Reeve N, Hobrath JV, Reynolds RC, Calderón AI. Identification of shikimate kinase inhibitors among anti-*Mycobacterium tuberculosis* compounds by LC-MS. *Tuberculosis (Edinb)*. 2014;94:152-188. doi:10.1016/j.tube.2013.12.004.

86. Rios-Soto L, Téllez-Valencia A, Sierra-Campos E, et al. Finding the first potential inhibitors of shikimate kinase from methicillin resistant *Staphylococcus aureus* through computer-assisted drug design. *Molecules*. 2021;26:6736. doi:10.3390/molecules26216736.

87. Javid-Majd F, Yang D, Ioerger TR, Sacchettini JC. The 1.25 Å resolution structure of phosphoribosyl-ATP pyrophosphohydrolase from *Mycobacterium tuberculosis*. *Acta Crystallogr Sect D Biol Crystallogr*. 2008;64:627-635. doi:10.1107/S0907444908007105.

88. McFeeters H, Gilbert MJ, Thompson RM, Setzer WN, Cruz-Vera LR, McFeeters RL. Inhibition of essential bacterial peptidyl-tRNA hydrolase activity by tropical plant extracts. *Nat Prod Commun*. 2012;77:1107-1110. doi:10.1177/1934578X1200700836.

89. Strange DS, Gaffin SS, Holloway WB, et al. Natural product inhibition and enzyme kinetics related to phylogenetic characterization for bacterial peptidyl-tRNA hydrolase 1. *Molecules*. 2021;26:2281. doi:10.3390/molecules26082281.

90. Ren J, Kotaka M, Lockyer M, Lamb HK, Hawkins AR, Stammers DK. GTP cyclohydrolase II structure and mechanism. *J Biol Chem*. 2005;280:36912-36919. doi:10.1074/jbc.M507725200.

91. Paxman JJ, Borg NA, Horne J, et al. The structure of the bacterial oxidoreductase enzyme DsbA in complex with a peptide reveals a basis for substrate specificity in the catalytic cycle of DsbA enzymes. *J Biol Chem*. 2009;284:17835-17845. doi:10.1074/jbc.M109.011502.

92. Mariano G, Monlezun L, Coulthurst SJ. Dual role for DsbA in attacking and targeted bacterial cells during type VI secretion system-mediated competition. *Cell Rep*. 2018;22:774-785. doi:10.1016/j.celrep.2017.12.075.

93. Lee Y, Kim Y, Yeom S, et al. The role of disulfide bond isomerase A (DsbA) of *Escherichia coli* O157:H7 in biofilm formation and virulence. *FEMS Microbiol Lett*. 2008;278:213-222. doi:10.1111/j.1574-6968.2007.00993.x.

94. Przepiora T, Figaj D, Bogucka A, et al. The periplasmic oxidoreductase DsbA is required for virulence of the phytopathogen *Dickeya solani*. *Int J Mol Sci*. 2022;23:697. doi:10.3390/ijms23020697.

95. Bocian-Ostrzycka KM, Grzeszczuk MJ, Banaś AM, Jagusztyn-Krynicka EK. Bacterial thiol oxidoreductases—from basic research to new antibacterial strategies. *Appl Microbiol Biotechnol*. 2017;101:3977-3989. doi:10.1007/s00253-017-8291-8.

96. Barra ALC, Dantas LOC, Morão LG, et al. Essential metabolic routes as a way to ESKAPE from antibiotic resistance. *Front Public Health*. 20;8. doi:10.3389/fpubh.2020.00026.

97. Khare G, Kar R, Tyagi AK. Identification of inhibitors against *Mycobacterium tuberculosis* thiamin phosphate synthase, an important target for the development of anti-TB drugs. *PLoS ONE*. 2011;6:e22441. doi:10.1371/journal.pone.0022441.

98. Mosbahi K, Wojnowska M, Albalat A, Walker D. Bacterial iron acquisition mediated by outer membrane translocation and cleavage of a host protein. *Proc Natl Acad Sci*. 2018;115:6840-6845. doi:10.1073/pnas.1800672115.

99. Skaar EP. The battle for iron between bacterial pathogens and their vertebrate hosts. *Plos Pathog*. 2010;6:1-2. doi:10.1371/journal.ppat.1000949.

100. Lemos ML, Balado M. Iron uptake mechanisms as key virulence factors in bacterial fish pathogens. *J Appl Microbiol*. 2020;129:104-115. doi:10.1111/jam.14595.

101. Khasheii B, Mahmoodi P, Mohammadzadeh A. Siderophores: importance in bacterial pathogenesis and applications in medicine and industry. *Microbiol Res*. 2021;250:126790. doi:10.1016/j.micres.2021.126790.

102. Schröder I, Johnson E, de Vries S. Microbial ferric iron reductases. *FEMS Microbiol Rev*. 2003;27:427-447. doi:10.1016/S0168-6445(03)00043-3.

103. Cain TJ, Smith AT. Ferric iron reductases and their contribution to unicellular ferrous iron uptake. *J Inorg Biochem*. 2021;218:111407. doi:10.1016/j.jinorgbio.2021.111407.

104. Choby JE, Skaar EP. Heme synthesis and acquisition in bacterial pathogens. *J Mol Biol*. 2016;428:3408-3428. doi:10.1016/j.jmb.2016.03.018.

105. Drew D, North RA, Nagarathinam K, Tanabe M. Structures and general transport mechanisms by the major facilitator superfamily (MFS). *Chem Rev*. 2021;121:5289-5335. doi:10.1021/acs.chemrev.0c00983.

106. Dos Santos SC, Teixeira MC, Dias PJ, Sá-Correia I. MFS transporters required for multidrug/multixenobiotic (MD/MX) resistance in the model yeast: understanding their physiological function through post-genomic approaches. *Front Physiol*. 2014;5:180. doi:10.3389/fphys.2014.00180.

107. Hassan KA, Liu Q, Elbourne LDH, et al. Pacing across the membrane: the novel PACE family of efflux pumps is widespread in Gram-negative pathogens. *Res Microbiol*. 2018;169:450-454. doi:10.1016/j.resmic.2018.01.001.

108. Vetting MW, de Carvalho LPS, Yu M, et al. Structure and functions of the GNAT superfamily of acetyltransferases. *Arch Biochem Biophys*. 2005;433:212-226. doi:10.1016/j.abb.2004.09.003.

109. Shirmast P, Ghafoori SM, Irwin RM, et al. Structural characterization of a GNAT family acetyltransferase from *Elizabethkingia anophelis* bound to acetyl-CoA reveals a new dimeric interface. *Sci Rep*. 2021;11:1274. doi:10.1038/s41598-020-79649-5.

110. Vimberg V, Zieglerová L, Buriánková K, Branny P, Balíková Novotná G. VanZ reduces the binding of lipoglycopeptide antibiotics to *Staphylococcus aureus* and *Streptococcus pneumoniae* cells. *Front Microbiol*. 2020;11. doi:10.3389/fmicb.2020.00566.

111. Sur VP, Mazumdar A, Vimberg V, et al. Specific inhibition of vanZ-mediated resistance to lipoglycopeptide antibiotics. *Int J Mol Sci*. 2021;23:97. doi:10.3390/ijms23010097.

112. Raghunath P. Roles of thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH) in *Vibrio parahaemolyticus*. *Front Microbiol*. 2015;5. doi:10.3389/fmicb.2014.00805.

113. Letchumanan V, Chan KG, Lee LH. *Vibrio parahaemolyticus*: a review on the pathogenesis, prevalence, and advance molecular identification techniques. *Front Microbiol*. 2014;55. doi:10.3389/fmicb.2014.00705.

114. Lorenz SC, Monday SR, Hoffmann M, Fischer M, Kase JA. Plasmids from Shiga toxin-producing *Escherichia coli* strains with rare enterohemolysin gene (ehxA) subtypes reveal pathogenicity potential and display a novel evolutionary path. *Appl Environ Microbiol*. 2016;82:6367-6377. doi:10.1128/AEM01839-16.

115. Hua Y, Zhang J, Jernberg C, et al. Molecular characterization of the enterohemolysin gene (ehxA) in clinical Shiga toxin-producing *Escherichia coli* isolates. *Toxins (Basel)*. 2021;13:71. doi:10.3390/toxins13010071.

116. Lorenz SC, Son I, Maounounen-Laasri A, Lin A, Fischer M, Kase JA. Prevalence of hemolysin genes and comparison of ehxA subtype patterns in Shiga toxin-producing *Escherichia coli* (STEC) and non-STEC strains from clinical, food, and animal sources. *Appl Environ Microbiol*. 2013;79:6301-6311. doi:10.1128/AEM02200-13.

117. Linhartová I, Bumba L, Mašín J, et al. RTX proteins: a highly diverse family secreted by a common mechanism. *FEMS Microbiol Rev*. 2010;34:1076-1112. doi:10.1111/j.1574-6976.2010.00231.x.

118. Ahmad JN, Sebo P. Bacterial RTX toxins and host immunity. *Curr Opin Infect Dis*. 2021;34:187-196. doi:10.1097/QCO.0000000000000726.

119. Pérez-Reytor D, Jaña V, Pavez L, Navarrete P, García K. Accessory toxins of *Vibrio* pathogens and their role in epithelial disruption during infection. *Front Microbiol*. 2018;9. doi:10.3389/fmicb.2018.02248.

120. Thanabalu T. Substrate-induced assembly of a contiguous channel for protein export from *Ecoli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *EMBO J*. 1998;17:6487-6496. doi:10.1093/emboj/17.22.6487.

121. Lenders MHH, Weidtkamp-Peters S, Kleinschrodt D, Jaeger K-E, Smits SHJ, Schmitt L. Directionality of substrate translocation of the hemolysin A Type I secretion system. *Sci Rep*. 2015;5:12470. doi:10.1038/srep12470.

122. Wandersman C, Delepelaire P. TolC, an *Escherichia coli* outer membrane protein required for hemolysin secretion. *Proc Natl Acad Sci USA*. 1990;87:4776-4780. doi:10.1073/pnas.87.12.4776.

123. Titball RW. Bacterial phospholipases C. *Microbiol Rev*. 1993;57:347-366. doi:10.1128/mr.57.2.347-366.1993.

124. Titball RW, Hunter SE, Martin KL, et al. Molecular cloning and nucleotide sequence of the alpha-toxin (phospholipase C) of *Clostridium perfringens*. *Infect Immun*. 1989;57:367-376. doi:10.1128/iai.57.2.367-376.1989.

125. Wilson SK, Knoll LJ. Patatin-like phospholipases in microbial infections with emerging roles in fatty acid metabolism and immune regulation by Apicomplexa. *Mol Microbiol*. 2018;107:34-46. doi:10.1111/mmi.13871.

126. Anderson DM, Sato H, Dirck AT, Feix JB, Frank DW. Ubiquitin activates patatin-like phospholipases from multiple bacterial species. *J Bacteriol*. 2015;197:529-541. doi:10.1128/JB02402-14.

127. Gxalo O, Digban TO, Igere BE, Olapade OA, Okoh AI, Nwodo UU. Virulence and antibiotic resistance characteristics of *Vibrio* isolates from rustic environmental freshwaters. *Front Cell Infect Microbiol*. 2021;11. doi:10.3389/fcimb.2021.732001.

128. Williams SG, Attridge SR, Manning PA. The transcriptional activator HIyU of *Vibrio cholerae*: nucleotide sequence and role in virulence gene expression. *Mol Microbiol*. 1993;9:751-760. doi:10.1111/j.1365-2958.1993.tb01735.x.

129. Meuskens I, Saragliadis A, Leo JC, Linke D. Type V secretion systems: an overview of passenger domain functions. *Front Microbiol*. 2019;10. doi:10.3389/fmicb.2019.01163.

130. Costa TRD, Felisberto-Rodrigues C, Meir A, et al. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol*. 2015;13:343-359. doi:10.1038/nrmicro3456.

131. Chung HY, Na EJ, Lee KH, et al. Complete genome sequence of *Vibrio parahaemolyticus* FORC_023 isolated from raw fish storage water. *Pathog Dis*. 2016;74:ftw032. doi:10.1093/femspd/ftw032.

132. Salomon D, Gonzalez H, Updegraff BL, Orth K. *Vibrio parahaemolyticus* type VI secretion system 1 is activated in marine conditions to target bacteria, and is differentially regulated from system 2. *PLoS ONE*. 2013;8:e61086. doi:10.1371/journal.pone.0061086.

133. Ling H, Saeidi N, Rasouliha BH, Chang MW. A predicted S-type pyocin shows a bactericidal activity against clinical *Pseudomonas aeruginosa* isolates through membrane damage. *FEBS Lett*. 2010;584:3354-3358. doi:10.1016/j.febslet.2010.06.021.

134. Salomon D, Kinch LN, Trudgian DC, et al. Marker for type VI secretion system effectors. *Proc Natl Acad Sci USA*. 2014;111:9271-9276. doi:10.1073/pnas.1406110111.

135. Thomas NA, Deng W, Puente JL, et al. CesT is a multi-effector chaperone and recruitment factor required for the efficient type III secretion of both LEE- and non-LEE-encoded effectors of enteropathogenic *Escherichia coli*. *Mol Microbiol*. 2005;57:1762-1779. doi:10.1111/j.1365-2958.2005.04802.x.

136. Castiblanco LF, Triplett LR, Sundin GW. Regulation of effector delivery by type III secretion chaperone proteins in *Erwinia amylovora*. *Front Microbiol*. 2018;9. doi:10.3389/fmicb.2018.00146.

137. Little DJ, Coombes BK. Molecular basis for CesT recognition of type III secretion effectors in enteropathogenic *Escherichia coli*. *PLoS Pathog*. 2018;14:e1007224. doi:10.1371/journal.ppat.1007224.

138. Strozen TG, Li G, Howard SP. YghG (GspS β) is a novel pilot protein required for localization of the GspS β type II secretion system secretin of enterotoxigenic *Escherichia coli*. *Infect Immun*. 2012;80:2608-2622. doi:10.1128/IAI06394-11.

139. Nivaskumar M, Francetic O. Type II secretion system: a magic beanstalk or a protein escalator. *Biochim Biophys Acta*. 2014;1843:1568-1577. doi:10.1016/j.bbamcr.2013.12.020.

140. Bunkute E, Cummins C, Crofts FJ, Bunce G, Nabney IT, Flower DR. PIP-DB: the protein isoelectric point database. *Bioinformatics*. 2015;31:295-296. doi:10.1093/bioinformatics/btu637.

141. Kozlowski LP. Proteome-pI: proteome isoelectric point database. *Nucleic Acids Res*. 2017;45:D1112-D1116. doi:10.1093/nar/gkw978.

142. Issaq HJ, Veenstra T. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. *Biotechniques*. 2008;44:697-700. doi:10.2144/000112823.

143. Ikai A. Thermostability and aliphatic index of globular proteins. *J Biochem*. 1980;88:1895-1898. doi:10.1093/oxfordjournals.jbchem.a133168.

144. Gamage DG, Gunaratne A, Periyannan GR, Russell TG. Applicability of instability index for in vitro protein stability prediction. *Protein Pept Lett*. 2019;26:339-347. doi:10.2174/0929866526666190228144219.

145. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 1982;157:105-132. doi:10.1016/0022-2836(82)90515-0.

146. Zhao G, London E. An amino acid "transmembrane tendency" scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: relationship to biological hydrophobicity. *Protein Sci*. 2006;15:1987-2001. doi:10.1110/ps.062286306.

147. Scott MS, Calafell SJ, Thomas DY, Hallett MT. Refining protein subcellular localization. *Plos Comput Biol*. 2005;11:e66. doi:10.1371/journal.pcbi.0010066.

148. Yao Y, Lv Y, Li L, et al. Protein sequence information extraction and subcellular localization prediction with gapped k-Mer method. *BMC Bioinformatics*. 2019;20:719. doi:10.1186/s12859-019-3232-4.

149. Jurgenson CT, Begley TP, Ealick SE. The structural and biochemical foundations of thiamin biosynthesis. *Annu Rev Biochem*. 2009;78:569-603. doi:10.1146/annurev.biochem.78.072407.102340.

150. Hasnain G, Frelin O, Roje S, et al. Identification and characterization of the missing pyrimidine reductase in the plant riboflavin biosynthesis pathway. *Plant Physiol*. 2012;161:48-56. doi:10.1104/pp.112.208488.

151. Bermingham A, Derrick JP. The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery. *BioEssays*. 2002;24:637-648. doi:10.1002/bies.10114.

152. Bayly AM, Macreadie IG. Cytotoxicity of dihydropteroate in *Saccharomyces cerevisiae*. *FEMS Microbiol Lett*. 2002;213:189-192. doi:10.1111/j.1574-6968.2002.tb11304.x.

153. Galperin MY. "Conserved hypothetical" proteins: prioritization of targets for experimental study. *Nucleic Acids Res*. 2004;32:5452-5463. doi:10.1093/nar/gkh885.

154. Ijaq J, Chandrasekharan M, Poddar R, Bethi N, Sundararajan VS. Annotation and curation of uncharacterized proteins—challenges. *Front Genet*. 2015;6. doi:10.3389/fgene.2015.00119.

155. Guiral M, Prunetti L, Aussignargues C, et al. The hyperthermophilic bacterium *Aquifex aeolicus*: from respiratory pathways to extremely resistant enzymes and biotechnological applications. *Adv Microb Physiol* 2012;61:125-194. doi:10.1016/B978-0-12-394423-8.00004-4.

156. Kamble AD, Singh H. Finding novel enzymes by in silico bioprospecting approach. In: *Value-Addition in Food Products and Processing Through Enzyme Technology*. Elsevier;2022:347-364. doi:10.1016/B978-0-323-89929-1.00028-7.

157. Bharat Siva Varma P, Adimulam YB, Kodukula S. In silico functional annotation of a hypothetical protein from *Staphylococcus aureus*. *J Infect Public Health*. 2015;88:526-532. doi:10.1016/j.jiph.2015.03.007.

158. Ijaq J, Chandra D, Ray MK, Jagannadham MV. Investigating the functional role of hypothetical proteins from an antarctic bacterium *Pseudomonas* sp. Lz4W: emphasis on identifying proteins involved in cold adaptation. *Front Genet*. 2022;13:825269. doi:10.3389/fgene.2022.825269.

159. Bjornson HS. Enzymes associated with the survival and virulence of gram-negative anaerobes. *Rev Infect Dis*. 1984;6(Suppl. 1):S21-S24. doi:10.1093/clinids/6.Supplement_1.S21.

160. Settembre E, Begley TP, Ealick SE. Structural biology of enzymes of the thiamin biosynthesis pathway. *Curr Opin Struct Biol*. 2003;13:739-747. doi:10.1016/j.sbi.2003.10.006.

161. Pohl M, Sprenger GA, Muller M. A new perspective on thiamine catalysis. *Curr Opin Biotechnol*. 2004;15:335-342. doi:10.1016/j.copbio.2004.06.002.

162. de Jong L, Meng Y, Dent J, Hekimi S. Thiamine pyrophosphate biosynthesis and transport in the nematode *Caenorhabditis elegans*: sequence data from this article have been deposited with the EMBL/GenBank data libraries under accession no. AY513235. *Genetics*. 2004;168:845-854. doi:10.1534/genetics.104.028605.

163. Kronenberger T, Schettert I, Wrenger C. Targeting the vitamin biosynthesis pathways for the treatment of malaria. *Future Med Chem*. 2013;55:769-779. doi:10.4155/fmc.13.43.

164. Müller IB, Hyde JE, Wrenger C. Vitamin B metabolism in *Plasmodium falciparum* as a source of drug targets. *Trends Parasitol*. 2010;26:35-43. doi:10.1016/j.pt.2009.10.006.

165. Backstrom AD, McMordie RAS, Begley TP. Biosynthesis of Thiamin I: the function of the thiE gene product. *J Am Chem Soc*. 1995;117:2351-2352. doi:10.1021/ja00113a025.

166. Lin Z, Xu Z, Li Y, Wang Z, Chen T, Zhao X. Metabolic engineering of *Escherichia coli* for the production of riboflavin. *Microb Cell Fact*. 2014;13:1-12. doi:10.1186/s12934-014-0104-5.

167. Eshik MME, Punom NJ, Begum MK, Khan T, Saha ML, Rahman MS. Molecular characterization of acute hepatopancreatic necrosis disease causing *Vibrio parahaemolyticus* strains in cultured shrimp *Penaeus monodon* in south-west farming region of Bangladesh, Dhaka Univ. *J Biol Sci*. 2018;27:57-68. doi:10.3329/dujbs.v27i1.46411.

168. Eshik MME, Abedin MM, Punom NJ, Begum MK, Rahman MS. Molecular identification of AHPND positive *Vibrio parahaemolyticus* causing an outbreak in south-west shrimp farming regions of Bangladesh. *J Bangladesh Acad Sci*. 2018;41:127-135. doi:10.3329/jbas.v41i2.35492.

169. Kumar S, He G, Kakarla P, et al. Bacterial multidrug efflux pumps of the major facilitator superfamily as targets for modulation. *Infect Disord Drug Targets*. 2016;16:28-43. doi:10.2174/1871526516666160407113848.

170. Favrot L, Blanchard JS, Vergnolle O. Bacterial GCN5-Related N-acetyltransferases: from resistance to regulation. *Biochemistry*. 2016;55:989-1002. doi:10.1021/acs.biochem.5b01269.

171. Assis PA, Espíndola MS, Paula-Silva FW, et al. *Mycobacterium tuberculosis* expressing phospholipase C subverts PGE2 synthesis and induces necrosis in alveolar macrophages. *BMC Microbiol*. 2014;14:128. doi:10.1186/1471-2180-14-128.

172. Monturiol-Gross L, Villalta-Romero F, Flores-Díaz M, Alape-Girón A. Bacterial phospholipases C with dual activity: phosphatidylcholinesterase and sphingomyelinase. *FEBS Open Bio*. 2021;11:3262-3275. doi:10.1002/2211-5463.13320.

173. Le Chevalier F, Cascioferro A, Frigui W, et al. Revisiting the role of phospholipases C in virulence and the lifecycle of *Mycobacterium tuberculosis*. *Sci Rep*. 2015;5:16918. doi:10.1038/srep16918.

174. Okino N, Ito M. Ceramidase enhances phospholipase C-induced hemolysis by *Pseudomonas aeruginosa*. *J Biol Chem*. 2007;282:6021-6030. doi:10.1074/jbc.M603088200.

175. Peterson JW. Bacterial pathogenesis. http://www.ncbi.nlm.nih.gov/pubmed/21413346. Published 1996.

176. Sharma AK, Dhasmana N, Dubey N, et al. Bacterial virulence factors: secreted for survival. *Indian J Microbiol*. 2017;57:1-10. doi:10.1007/s12088-016-0625-1.

177. Fleitas Martínez O, Cardoso MH, Ribeiro SM, Franco OL. Recent advances in anti-virulence therapeutic strategies with a focus on dismantling bacterial membrane microdomains, toxin neutralization, quorum-sensing interference and biofilm inhibition. *Front Cell Infect Microbiol.* 2019;9:74. doi:10.3389/fcimb.2019.00074.

178. Granato ET, Harrison F, Kümmerli R, Ross-Gillespie A. Do bacterial "virulence factors" always increase virulence? A meta-analysis of pyoverdine production in *Pseudomonas aeruginosa* as a test case. *Front Microbiol.* 2016;7:1952. doi:10.3389/fmicb.2016.01952.

179. Speer AL, Grikscheit TC, Upperman JS, Ford HR. Sepsis and related considerations. In: *Pediatric Surgery*. Elsevier;2012:141-163. doi:10.1016/B978-0-323-07255-7.00010-6.

180. Kane TL, Carothers KE, Lee SW. Virulence factor targeting of the bacterial pathogen *Staphylococcus aureus* for vaccine and therapeutics. *Curr Drug Targets.* 2018;19:111-127. doi:10.2174/1389450117666161128123536.

181. Andersson JA, Sha J, Erova TE, et al. Identification of new virulence factors and vaccine candidates for *Yersinia pestis. Front Cell Infect Microbiol.* 2017;7:448. doi:10.3389/fcimb.2017.00448.