Research Article

Questing functions and structures of hypothetical proteins from *Campylobacter jejuni*: a computer-aided approach

Md. Amran Gazi[1], Sultan Mahmud[2], Shah Mohammad Fahim[1], Md. Rezaul Islam[3], Subhasish Das[1], Mustafa Mahfuz[1,4] and Tahmeed Ahmed[1]

[1]Nutrition and Clinical Services Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Bangladesh; [2]Infectious Diseases Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Bangladesh; [3]Neurosciences and Molecular Biosciences, International Max Planck Research School, Grisebachstraße 5, 37077 Göttingen, Germany; [4]Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland

Correspondence: Md. Amran Gazi (amran.gazi@icddrb.org)

Campylobacter jejuni (*C. jejuni*) is considered to be one of the most frequent causes of bacterial gastroenteritis globally, especially in young children. The genome of *C. jejuni* contains many proteins with unknown functions termed as hypothetical proteins (HPs). These proteins might have essential biological role to show the full spectrum of this bacterium. Hence, our study aimed to determine the functions of HPs, pertaining to the genome of *C. jejuni*. An *in-silico* work flow integrating various tools were performed for functional assignment, three-dimensional structure determination, domain architecture predictors, subcellular localization, physicochemical characterization, and protein–protein interactions (PPIs). Sequences of 267 HPs of *C. jejuni* were analyzed and successfully attributed the function of 49 HPs with higher confidence. Here, we found proteins with enzymatic activity, transporters, binding and regulatory proteins as well as proteins with biotechnological interest. Assessment of the performance of various tools used in this analysis revealed an accuracy of 95% using receiver operating characteristic (ROC) curve analysis. Functional and structural predictions and the results from ROC analyses provided the validity of *in-silico* tools used in the present study. The approach used for this analysis leads us to assign the function of unknown proteins and relate them with the functions that have already been described in previous literature.

Introduction

Campylobacter is the genus that comprises a diverse group of non-spore forming rod-like or spiral-shaped Gram-negative bacteria [1]. In developing countries, infections with *Campylobacter* are common in children under 2 years of age and found to be associated with increased incidence of diarrheal diseases as well as mortality [1,2]. In industrialized nations, *Campylobacter* is the cause of diarrhea during early years of adulthood [3]. *Campylobacter* infections are mostly acquired through consumption of contaminated water and food in resource-poor environment [4]. Two of the species, *C. jejuni* and *C. coli*, are primarily known to be responsible for human campylobacteriosis [4]. Acute gastroenteritis and food poisoning can be induced by *C. jejuni* in infected patients. Usually, *C. jejuni* infection causes gastroenteritis without any complication but acute infection may results in abdominal cramps, fever or other ailments like Guillain–Barré syndrome or Miller Fisher syndrome [5]. Recent studies also showed an association of *Campylobacter* infections with malnutrition, a condition highly prevalent in developing countries [2].

1

Although whole genome sequence of *C. jejuni* NCTC has been published, a detailed catalog of prospective virulence is yet to be documented. Its complete genome contains a circular chromosome of 1641481 base pairs with GC content: 30.6%. Several studies since then suggest *C. jejuni* exhibits high genomic diversity across strains. A shotgun DNA microRNA approach revealed 63-kb long unique genomic DNA sequences in another *Campylobacter* strain, *C. jejuni* 81–176 when compared with fully sequenced *C. jejuni* NCTC 11168, implying genetic diversity between strains [6,7]. Overall, genome of *C. jejuni* strain 81–176 (total length 1.6 Mb) available in NCBI encodes 1658 proteins (GC%: 30.4) [7]. Among them 267 are yet to be experimentally determined, and are designated as hypothetical proteins (HPs). Similar to functionally annotated proteins, HP originates from an open reading frame (ORF), but lacks functional annotations [8]. Therefore, annotation of HPs of specific organism leads to the introduction of unique functions, and helps in listing auxiliary protein pathways [8].

Several contemporary bioinformatics tools, for instance, CDART, SMART, Pfam, INTERPROSCAN, MOTIF, SUPERFAMILY, and SVMProt have been well established to specify the functions of many bacterial HPs [9–11]. Besides, the exploration of protein–protein interaction (PPI) for instance, using STRING database [12], is crucial for comprehending the aspect of biological network. During cellular processes protein interactions play an essential role. Thus, an understanding of HP function can be reached by studying the PPIs [13]. Consequently, interaction of one protein and their function is proven to be dependent on the regulatory connection with other protein [54]. Three-dimensional modeling is also a great way to relate structural knowledge with the function of undetermined proteins [14]. Protein structure is generally more conserved than protein sequence [15]. Therefore, structural determination is considered to be a strong indicator of similar function in two or more proteins. Moreover, evolutionary distant proteins and its function can also be identified through structural information [15].

Functional prediction of HPs by using *in silico* approaches has been successfully applied for various bacteria and parasites [10,16,17]. In the present study, we have chosen *C. jejuni* as a template to explore the functions of HPs from its genome with a higher accuracy using well-optimized bioinformatics tools.

Materials and methods
Retrieval of genome data

Full genome of *C. jejuni* strain 81–176 was retrieved from NCBI (GCA_000015525.1, NC_008787.1). According to the repository this genome encodes 1658 proteins (http://www.ncbi.nlm.nih.gov/genome/), of which 267 are assigned as HPs. FASTA sequences of HPs were then retrieved for further analysis in the present study (accessed 27 February 2019).

Functional analysis of HPs

In order to assign the function using the databases depicted in Supplementary Table S1, first we submitted proteins to five publicly available free tools (CDD-BLAST, HmmScan, SMART, Pfam, and SCANPROSITE) [18–22]. These databases can search for the conserved domains and subsequently help in the categorization of proteins. Analyses of HPs by five webtools revealed the distinct results. To find a composite result, different confidence levels were assigned on the basis of pooled results obtained from five webtools. For instance, if we observed same results from the five distinct tools, the composite score was 100 (percentage of confidence). For downstream analyses, we filtered 50 out of 267 HPs that displayed 60% or above confidence (Supplementary Table S2).

Next, we performed functional assignment of these 50 selected HPs using different tools (Figure 1). SMART and CDART [23] facilitated to look for functions using the domain architecture and conserved domain database, respectively. To classify HPs into functional families based on similarity, we employed SUPERFAMILY [24], Pfam [21], and SVMProt [25]. Software such as InterPro and MOTIF search tool were also used to detect the motif in the proteins [26,27]. Default parameters were used for all these databases.

We further annotated HPs manually through searching for homologous proteins from related organisms. To do this, we used BLAST against the NCBI nonredundant (nr) database. If the two sequences were ≥90% identical, we considered it as homologues to each other. Query cover, score parameters and e-value of every hit are summarized in Supplementary Material S5.

Geptop 2.0 database was used to identify the essential genes among the HPs [28]. Default essentiality score cutoff of 0.24 was adopted. Geptop is the essential gene identification tool based on phylogeny and orthology. In the present study, a similarity search was also done against DrugBank 3.0 for all the targets [29].
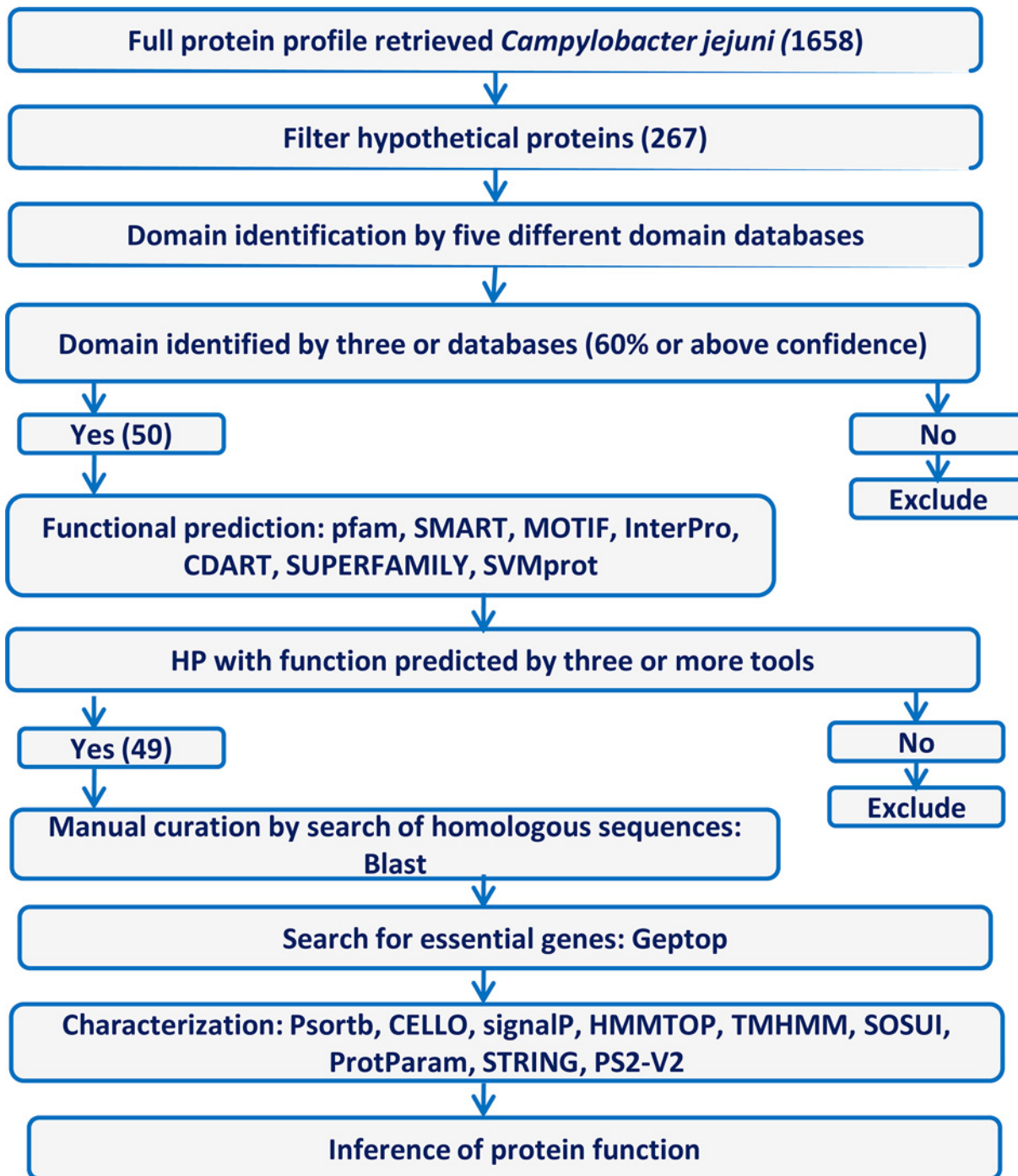
Figure 1. Flow chart showing the overall design of the study

Prediction of physicochemical characteristics

Expasy's ProtParam server was used for extinction coefficient, isoelectric point (pI), molecular mass, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) prediction [30].

Identification of subcellular localization

PSORTb [31] and CELLO [32] were applied to find the localization of HPs in the cell. PSORTb contains the information both from laboratory experimentations and *in silico* prediction. On the other hand, a support vector machine

3

was used by CELLO database to generate the probable localization of protein in the cell. TMHMM [33], SOSUI [34], HMMTOP [35], and SignalP [36] were also applied to detect membrane protein and to verify the presence of cleavage sites for peptide.

Functional protein association networks

We had employed STRING software [37] to predict interactive partners of HPs in this investigation. This database computes the network based on physical and functional associations. Highest score network proteins were selected for this analysis in order to accord the reliability of the PPIs.

Determination of three-dimensional structures

Structure prediction of a protein from its sequences is a way that enables the identification of function. A template based online server PS2-v2 was used to predict the tertiary structure of the HPs in this study [38]. This server uses a template of known protein structures and then applied the approaches of multiple and pairwise alignments combining IMPALA, T-COFFEE, and PSI-BLAST.

Performance assessment

A receiver operating characteristic (ROC) was implemented to confirm the accuracy of the predicted functions of HPs from *C. jejuni* genome. First, we selected 40 proteins randomly with known functions of *C. jejuni* (Supplementary Table S3). These proteins were predicted for the functions using the same databases that were used for the prediction of HPs. To classify the prediction, true positive (1) and true negative (0) were denoted as binary numerals. Six levels diagnostic efficacy was also evaluated where the integers '2', '3', '4', and '5' were used. A web-based calculator was applied to submit the classification data for ROC curve and is utilized to calculate the sensitivity, specificity, ROC area, and accuracy of the tools used to speculate the function of HPs [39].

Results and discussion
Analysis of HPs from *C. jejuni* genome

With the ongoing developments of DNA sequencing technologies called high throughput sequencing techniques has enabled a substantial number of bacterial genome sequencing. Annotation of the genes generally depends on sequence homology techniques [40]. However, a large number of genes have no assigned function. Therefore, only homology techniques cannot assign functions precisely and may lead to incorrect annotations [41]. Multiple tools should be used to avoid this problem to assign functions of HPs. Hence, the present study focused on the annotation of HPs from *C. jejuni* using assorted but effective bioinformatics tools.

First, functional domains were identified from the sequences of all the 267 HPs using SCANPROSITE, SMART, Pfam, CDD-BLAST, and HmmScan. Specific domains could be identified using one, two, three, four, or five of the above-stated tools and therefore, different confidence levels were assigned (e.g., 20, 40, 60, 80, and 100%). In our previous studies, published elsewhere, we only considered the proteins with 100% confidence [10,42]. However, in the current study, HPs having 60% or above confidence level have been considered to gain the greater coverage. The analyses revealed 50 such proteins which were used for downstream analyses. For rest of the HPs (n=217), domains were recognized from one or two of the mentioned tools. Further studies are needed to find the exact function for these proteins. Supplementary Table S2 summarized protein lists with domain. The final pool of 50 proteins was examined employing CDD-BLAST, Pfam, SMART, MOTIF, InterPro, CDART, SUPERFAMILY, and SVMProt. Functional annotation was considered to be high for proteins that manifested same function from equal or more than three tools (Supplementary Table S4). Thus, we inferred 49 such proteins with high confidence (Table 1) and classified them as highly confident proteins (Hconf), where 11 contain homologous sequences without product function reported (Supplementary Table S5). Analyses of sequence were then accumulated and Hconf proteins were grouped into different functional categories. Functional classes of proteins consists of regulatory proteins, transporters, binding proteins, enzymes, proteins with biotechnological interest, and proteins with other functions (Figure 2). The categorization was selected based on the literature search and gene ontology. Enzyme classes were determined from enzyme data bank of Expasy (https://enzyme.expasy.org/cgi-bin/enzyme/enzyme-search-cl?2).

Moreover, essential genes were predicted using Geptop, a database that accommodates already sequenced bacterial genomes. These genes are fundamental for survival of an organism and perform essential activities of the cell [43]. Identification of essential genes is an important stride toward gaining better insight into the evolution [44]. Time-absorbing and challenging experiential procedures like transposon mutagenesis, RNA interference, and single-gene knockouts were used to identify essential genes [28]. However, *in-silico* approaches offer an alternative

Table 1 HPs functionally annotated from *C. jejuni*

| No.	Protein IDs	Protein function
1	WP_002868767.1	Curli production assembly, transport component CsgG
2	WP_002854524.1	Chemotaxis phosphatase CheX
3	WP_009882162.1	SprA-related family
4	WP_010790856.1	Pyridoxamine 5′-phosphate oxidase
5	WP_009882239.1	Hemagglutination activity domain
6	WP_002854991.1	FxsA cytoplasmic membrane protein, FxsA
7	WP_002855029.1	DNA replication regulator, HobA
8	WP_002868905.1	GDSL-like lipase
9	WP_002869356.1	Divergent polysaccharide deacetylase
10	WP_002856929.1	C4-type zinc ribbon domain
11	WP_002869028.1	Esterase-like activity of phytase
12	WP_011812736.1	Domain of unknown function DUF234
13	WP_002868809.1	Ankyrin repeats, Ank_2
14	WP_002869368.1	Type-1V conjugative transfer system mating pair stabilization, TraN
15	WP_009882583.1	NLPC_P60 stabilizing domain
16	WP_002853389.1	Jag, N-terminal domain superfamily
17	WP_009882608.1	Adhesin from *Campylobacter*
18	WP_002856369.1	Putative β-lactamase-inhibitor-like
19	WP_079254190.1	β-1,4-N-acetylgalactosaminyltransferase (CgtA)
20	WP_002856180.1	Heavy-metal-associated domain
21	WP_002831611.1	Transcription factor zinc-finger
22	WP_002790076.1	Methyl-accepting chemotaxis protein (MCP) signaling domain
23	WP_002853792.1	Plasminogen-binding protein pgbA N-terminal
24	WP_002869072.1	Putative S-adenosyl-L-methionine-dependent methyltransferase
25	WP_002869097.1	MaoC-like dehydratase domain
26	WP_002869326.1	Metallo-carboxypeptidase
27	WP_002869139.1	Pyruvate phosphate dikinase, PEP
28	WP_002869195.1	Anti-sigma-28 factor
29	WP_002856630.1	PD-(D/E)XK nuclease superfamily
30	WP_002855458.1	MgtE intracellular N domain
31	WP_002797496.1	Flagellar FliJ protein
32	WP_024088174.1	Nitrate reductase chaperone
33	WP_009883030.1	ATPase, AAA-type, core
34	WP_002824979.1	Putative NADH-ubiquinone oxidoreductase chain E
35	WP_002869225.1	DMSO reductase anchor subunit (DmsC)
36	WP_002856602.1	Putative β-lactamase-inhibitor-like
37	WP_002868888.1	Tetratricopeptide repeat, TPR_2
38	WP_002868880.1	ABC-type transport auxiliary lipoprotein component
39	WP_009883121.1	Flagellar FLiS export co-chaperone
40	WP_002860117.1	Menaquinone biosynthesis
41	WP_002779704.1	T-antigen specific domain
42	WP_011187233.1	Toprim domain
43	WP_011187235.1	AAA domain, AAA_25
44	WP_002809111.1	TrbM superfamily
45	WP_011117548.1	Bacterial virulence protein VirB8
46	WP_011117549.1	Conjugal transfer protein
47	WP_011117575.1	Type IV secretion system proteins,T4SS
48	WP_011799393.1	TrbM superfamily
49	WP_011117588.1	mRNA interferase PemK-like
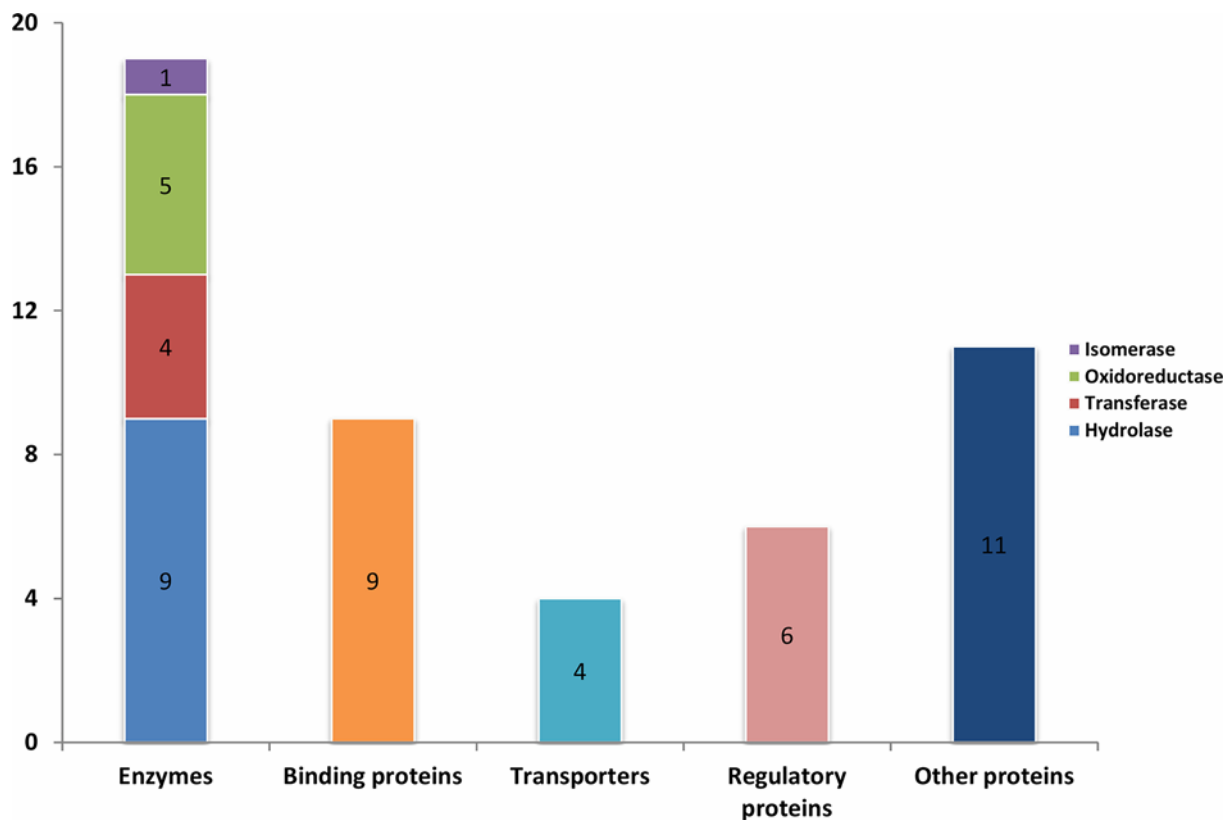
Figure 2. Functional classification of 49 HPs into various groups

Table 2 ROC results of various tools used in the present study

No.	Software	Accuracy (%)	Sensitivity (%)	Specificity (%)	ROC area
1	PFAM	95%	94.7%	100%	0.97
2	SMART	95%	94.9%	100%	0.97
3	MOTIF	95%	94.9%	100%	0.97
4	INTERPROSCAN	95%	94.9%	100%	0.97
5	CDART	97.5%	97.4%	100%	0.99
6	SUPERFAMILY	95%	94.1%	100%	0.97
7	SVMprot	90%	88.9%	100%	0.94
8	Average	**95%**	**94.3%**	**100%**	**0.97**

for predicting essential genes. In the current study, it was possible to identify 32 essential proteins by using Gep-top database (Supplementary Table S6). Besides, from the selected Hconf proteins, only one protein was found to be exhibited similarity with approved drugs. The test was done through protein BLAST against DrugBank. Protein WP_002868809.1 showed the similarity with fostamatinib that could act as inhibitors. DrugBank contains 6816 FDA-approved and experimental drugs, 169 drug enzymes/carriers, and 4326 drug targets.

Finally, ROC curve was calculated to identify the reliability of the tools used to predict the function. Average accuracy was found to be 95% for the used pipeline and area under the curve (AUC) was 0.97 (Table 2). It is recommended to use the AUC to summarize the overall accuracy of the tools in the diagnosis [45]. The AUC value ranges from 0 to 1, and the value greater than 0.7 is considered acceptable [45]. The ROC analyses results provided the high reliability of *in-silico* tools used in our study (Table 2). However, predicting the functions of the 'function-known' proteins and obtaining very high accuracy does not mean the prediction on 'function-unknown' proteins would reproduce the same level of accuracy.

Enzymes

We found five oxidoreductases among these HPs of *C. jejuni*. These enzymes play key role in the pathogenesis. WP_002824979.1 is an NADH-quinone oxidoreductase, an enzyme that involves in regulating the expression of virulence factors, electron transport, and sodium translocation [46]. This putative domain commonly found in Epsilonproteobacteria, chiefly in *Helicobacter pylori* (*H. pylori*) [47]. Protein WP_002869225.1 is dimethyl sulfoxide reductase that acts as the terminal electron transfer enzyme in *Escherichia coli* (*E. coli*). This enzyme and the reaction it catalyzes could prove helpful on the climate control frontier [48]. We also found four proteins as transferase those might involved in bacterial pathogenesis and virulence. Among them, protein WP_002854524.1 is responsible for modifying the bacterial character in the presence of repellents and nutrients, found in chemotaxis phosphatase CheX [49]. Hydrolases is the third class of enzymes where almost 50% proteins among all characterized enzymes represent this class. This class of proteins is generally membrane-bound involved in various virulence factors associated with metal ion binding, transmembrane transport, cell wall degradation. We have found WP_002856630.1 that represents endonuclease-like domain involved in DNA repair and replication [50]. WP_009883030.1 and WP_011187235.1 exhibit AAA ATPases (ATPases associated with diverse cellular activities) which plays a number of role in the cell including protein proteolysis and disaggregation, cell-cycle regulation, organelle biogenesis, and intracellular transport [51]. In addition WP_011187233.1 protein is a toprim (topoisomerase-primase) domain that is found in bacterial DnaG-type primases, involved in DNA strand breakage and rejoining [52].

Binding

We have identified nine proteins as binding among the functionally annotated HPs. These can be further classified into RNA binding, DNA binding, protein binding, ion binding, and adhesion proteins. Binding of proteins is important in the propagation and survival of pathogens in the host [53]. For example, protein binding WP_002868888.1 is tetratricopeptide repeat (TPR) motifs, reported to be directly related to virulence-associated functions [54]. WP_002853792.1 is the N-terminal domain of the bacterial proteins (PgbA) that bind to host cell protein, plasminogen [55]. This activity was identified in *H. pylori* where it is thought to contribute to the virulence of this bacterium [55]. WP_011117588.1 is mRNA interferase PemK-like domain, a growth inhibitor in *E. coli*. It is responsible for mediating cell death through inhibiting protein synthesis [56]. Besides, WP_009882239.1 is a hemagglutination activity domain found in a number of large, repetitive proteins of bacteria. Filamentous hemagglutinin (FHA) is a secreted and surface-exposed protein that acts as main virulence attachment factor in childhood whooping cough caused by *Bordetella pertussis* [57]. WP_002868809.1 is found to be ankyrin repeat (ANK), a typical PPI motif in nature. A large number of bacterial pathogens mimic or manipulate various host functions through delivering ANK-containing proteins into eukaryotic cells [58]. Finally, WP_009882608.1 is adhesion protein called surface-exposed lipoprotein JlpA, an early critical step in the pathogenesis of *C. jejuni* disease [59]. This HP might provide new approach for the rational design of small molecule inhibitors against *C. jejuni* targeting JlpA efficiently [59].

Regulatory

There are six HPs found to be involved in regulatory and cellular mechanisms, and are essential for the pathogenesis of *C. jejuni*, hence can be treated as probable drug targets. WP_002869195.1 is found to be anti-sigma-28 factor that inhibits the activity of the sigma 28 transcription factor. This inhibition prevents the expression of genes from flagellar transcriptional class 3, which include genes for chemotaxis. Mechanism of action of anti-sigma factors has opened new door on the regulation of bacterial gene expression, as anti-sigma factors join another layer to transcriptional control via negative regulation. The bacteriophage T_4 uses an anti-sigma factor in order to transcribe its own genes by sabotaging the *E. coli* RNA polymerase [60]. WP_002797496.1 is a membrane-associated protein that affects chemotactic events. FliJ is a component of the flagellar export and has a chaperone-like activity. Mutations in FliJ result in failure to respond to chemotactic stimuli [61]. Moreover, WP_011117549.1 is identified as conjugal transfer protein that bacteria utilize to export effector molecules during infection. For example, *H. pylori* use type IV machines to transport effectors to the extracellular environment or cell cytosol of mammals [62]. A DnaA binding protein (WP_002855029.1) HobA, identified that is an essential regulator of DNA replication in *H. pylori* [63]. WP_002790076.1 is methyl-accepting chemotaxis protein (MCP) that allows bacteria to sense the concentrations of molecules (nutrients/toxins) in the extracellular milieu so that they can smooth swim or fall accordingly [64].

Transporters

Transporter proteins are involved various metabolic processes, are responsible for transportation of nutrients, and hence, essential for survival of the organism. Besides, they accelerate the movement of virulence factors and are directly involved in pathogenesis [65]. WP_002855458.1 is the magnesium transporter E (MgtE), found in eukaryotic proteins. Magnesium (Mg^{2+}) is an essential element for growth and maintenance of living cells where MgtE transports magnesium across the cell membrane [66]. WP_002868880.1 is an ABC-type transport, responsible for outer membrane biosynthesis in bacteria that can be an excellent drug target [67]. WP_002856180.1 is heavy metal-associated (HMA) domain found in a number of detoxification proteins or in heavy metals transport. Proteins that are involved in transporting heavy metals in bacteria, plants, and mammals share similarities across the kingdoms in their structures and sequences. These proteins provide an important arena for research, some being involved in bacterial resistance to toxic metals, while others are responsible for acquired human diseases, such as Wilson's and Menke's diseases [68]. WP_011117548.1 is the bacterial virulence protein VirB8 that is thought to be a constituent of DNA transporter. In addition, VirB8 is a potential drug target that targets its PPIs. X-ray structure has enabled a detailed structure–function analysis of VirB8, which identified VirB8 interaction with VirB4 and VirB10 [69]. Our results also go in line with this as we observed VirB8 has strong interaction with VirB10.

Potential proteins with biotechnological applications

We identified few proteins that can have biotechnological applications based on their functional process. For instance, WP_010790856.1 is pyridoxamine 5′-phosphate oxidase (pdxH), an enzyme involved in the *de novo* synthesis of pyridoxal phosphate and pyridoxine (vitamin B6). Moreover, PdxH is evolutionarily related to phzD (also known as phzG), one of the enzymes in the phenazine biosynthesis protein pathway [70]. Only known source of phenazines are bacteria in nature. This is used as drug and also acts as biocontrol agents to inhibit plant pests. For example, the phenazine pyocyanin contributes to its potential to colonize the lungs of cystic fibrosis patients [71]. Similarly, phenazine-1-carboxylic acid, produced by a number of *Pseudomonas*, increases survival in soil and has been shown to be important for the biological control of certain strains [72]. The protein WP_002869072.1 was predicted to be S-adenosyl-L-methionine-dependent methyltransferase (SAM-MTase). Methyltransferases transfer a methyl group from a donor to an acceptor during methylation of biopolymers [73]. SAM-MT was used in the pharmaceutical industry as catechol, first as an antimicrobial and anticancer agent [73,74].

Protein WP_024088174.1 is the nitrate reductase that produces nitrite from nitrate. Nitrate is the primary source of nitrogen in fertilized soils and the reaction is critical for the production of protein in crop plants. Nitrate reductase enzyme activity can also be used as a biochemical tool for predicting grain protein production and subsequent grain yield. For example, it promotes amino acid content in tea leaves [75]. It is also reported that tea plants sprayed with various micronutrients (like Zn, Mn, and B) along with Mo enhanced the amino acid production of tea and the crop yield [75]. WP_002869028.1 is a phytase-like domain that catalyzes the hydrolysis of phytic acid. Phytic acid is organic form of phosphorus and indigestible found in grains and oil seeds. Phytase is produced by bacteria found in the gut of ruminant animals which are able to make phosphorus from phytic acid [76]. But, non-ruminants like human cannot make phytase. Research in the field of animal nutrition has put the idea of supplementing feed with phytase to make sure the availability of phytate-bound nutrients like phosphorus, calcium, carbohydrates, proteins, and other minerals [77].

Peptidase, an enzyme that is used as the ingredients of detergents, foods, and pharmaceuticals [78]. In this study, WP_009882583.1 was found to be cysteine peptidase that hydrolyzes a peptide bond utilizing the thiol group of cysteine as nucleophile. These peptidases are often confined to acidic environments and active at acidic pH such as the plant vacuole or animal lysosome. WP_002868905.1 is GDSL esterases and lipases are hydrolytic enzymes with broad substrate specificity. They have potential for use in the synthesis and hydrolysis ester compounds of biochemical, food, pharmaceutical, and other biological interests [79].

Other proteins

WP_002856369.1 and WP_002856602.1 was found to be β-lactamase-inhibitor, a group of enzymes responsible for bacterial resistance to β-lactam antibiotics [80]. WP_009883121.1 s ass fla agellar FLiS export co-chaperone. Previously, various FliS-associated proteins in *H. pylori* were identified by a yeast two-hybrid study, but the implications are unknown [81]. Chaperones are usually involved in various important processes such as protein degradation, folding, and polypeptide translocation [81].

At last, WP_002860117.1 protein family includes two enzymes involved in menaquinone (vitamin K2) biosynthesis. In prokaryotes, vitamin K2 serves as the sole quinone molecule in electron shuffling systems while menaquinone

pathway is absent from humans [82]. Therefore, novel antibacterial agents are possible to develop by targeting the bacterial enzymes responsible for menaquinone biosynthesis. It has been reported that inhibition of menaquinone showed significant growth inhibition against multidrug-resistant *Mycobacterium* and other Gram-positive bacteria as well as effective in killing Gram-negative bacteria [83].

Prediction of primary properties and protein localization

Sequences of amino acids of 49 HPs were analyzed to evaluate their primary properties, and their localization (Supplementary Table S7). But, we paid attention to some proteins that showed functions important for the survival of *Campylobacter* and might have biotechnological interest. The proteins WP_024088174.1, WP_002869072.1, WP_010790856.1, WP_002868905.1, WP_002869028.1, WP_009882583.1 all had molecular weight (MW) values between 15792.47 and 52423.83. These proteins are referred to be biotechnologically important in the present study. Some proteins, essential for pathogenesis of *Campylobacter* have MW ranged from 8773.25 to 39113.6. The pI is the pH where protein carries no net electrical charge. For the list of mentioned proteins, it ranged from 5.03 to 9.63.

The aliphatic index indicates the protein thermostability [84]. Protein WP_002856369.1, associated with β-lactamase inhibition showed the highest values of 133.14. The GRAVY of protein indicates its hydrophobicity or the interaction with water [85]. In WP_002869028.1, WP_009882583.1, and WP_024088174.1, the scores are among −0.744, −0.439, and −0.393. Moreover, the instability index offers an assumption of the stability of protein *in vitro*. We used cut-off values >40 and <40 to discriminate between stable and unstable proteins, respectively. From our listed proteins, WP_024088174.1 and WP_002868880.1 were considered to be stable.

Localization plays an essential role in determining function of unknown proteins [11]. Protein WP_002868905.1 and WP_009882583.1 is located in outer membrane whereas other proteins of interest were predicted to be in the cytoplasm.

PPI network

Function of a completely unknown protein can be identified based on the evidence of their interactions with the known proteins of a particular organism [11]. For example, PPI map and *in-vitro* proteome-wide interaction screens were applied to successfully assign the function of 50 unknown proteins for *Streptococcus pneumoniae* [86]. In our study protein WP_010790856.1, an oxidase (pdxH) showed a strong interaction with the Pyridoxine 5′-phosphate synthase that involved in vitamin B6 synthesis. WP_024088174.1 is interacted with formate dehydrogenase, an oxidoreductase that oxidizes formate to form carbon dioxide. WP_002868880.1 was found to be interacted with ABC transporter that functions to maintain the asymmetry of the outer membrane. All these predictions of functional partners have strengthened our findings of function predicted by using functional prediction tools (Supplementary Table S8).

Three-dimensional structures

Structural genomics has become a robust way to determine the novel structures of proteins, especially via X-ray crystallography [87]. Determination of unannotated protein structures can often help us to discover unexpected family relationships, hence giving the idea of their probable functions. Proteins unrelated to existing PDB entries may represent new functions. In this case, structures homologous to other organisms have manifested as surrogates in drug discovery. For example, Nolatrexed, an anticancer drug was discovered using the structure of *E. coli* thymidylate synthase (46% sequence identity with human homolog) [87]. Kinase inhibitors to kill the *Plasmodium falciparum* were identified using structures of protein kinases from *Cryptosporidium* and *Toxoplasma* (61 and 74% sequence identity, respectively) [88].

In our study, PS2-v2 online server was used to model the three-dimensional structures of the Hconf proteins for *Campylobacter*. Among the 49 Hconf proteins, 24 proteins revealed same domain as function prediction tools used in the present study. In contrast, nine proteins showed discrepant results and no suitable templates were found for 16 proteins (Supplementary Table S9). Identity of model ranged from 54.5 to 91.6% and was constructed from closely related *Campylobacter* genus bacteria belonging to the *H. pylori, E. coli, Bacillus*, and *Clostridium*.

Based on the resolution and identity, two best models were WP_002797496.1 and WP_002854991.1, which were annotated as Flagellar FliJ protein and FxsA cytoplasmic membrane protein, respectively. The structure obtained for FliJ protein was determined by X-ray crystallography earlier and refined with diffraction data to 1.8-Å resolutions, which was solved by an ortholog isolated from *Saccharomyces cerevisiae* (PDB 2efrA). FxsA was determined by electron microscopy and refined with diffraction data to 4-Å resolutions and solved by an ortholog isolated from

Torpedo marmorata (PDB 1oedB). Both these proteins showed the same function as predicted by other function prediction tools. Proteins with shared sequence typically display similar functions in this way.

Conclusions

Protein function identification of a pathogen is an essential step to understand its cellular and molecular processes. In the present study, we used a computer-aided approach to assign the function of HPs from *C. jejuni*. We predicted the function to 49 HPs with a higher confidence. In addition, localization of protein and primary structure prediction were useful in supporting the specific characteristics of annotated proteins. Proteins were further explored for PPI and their tertiary structures. We have identified proteins with important functions including enzymes, transporters, binding and regulatory proteins as well as proteins with biotechnological interest. To summarize, our comprehensive analysis produces a better understanding of *C. jejuni* genome related HPs that would help to find novel therapeutic interventions and targets. Moreover, we have obtained an excellent result using the pipeline used in the present study and the method can be used to annotate the function of unknown proteins.

However, biochemical and clinical investigations are required to confirm the function of predicted proteins. Several studies have been conducted previously using the cumulative *in-silico* and *in-vitro/in-vivo* approach to investigate the function of unknown proteins. For instance, *in silico* approaches were used to predict the biological function of some of the unknown *Mycobacterium* proteins. The chosen proteins posses the α/β- hydrolase topological fold, characteristic of lipases/esterases which were further validated by wet lab experiments [89]. Combination of *in-silico and in-vitro/in-vivo* assays were also used to characterize the function of HPs from several other organisms [90–93]. Moreover, *in-silico* structure prediction methods were applied for drug discovery in the absence of x-ray structure of the target protein and again confirmed by *in-vitro* assays. Nonetheless, functional prediction merely on *in silico* methods requires careful integration of several computational tools into a single streamlined process. We hope that the information of HPs in the present study will be innovative for further *in-vitro/in-vivo* analysis on *C. jejuni*.

Abbreviations

ABC , ATP-binding cassette; ANK , ankyrin repeat; AUC, area under the curve; FDA , food and drug administration; GDSL , motif consensus amino acid sequence of Gly, Asp, Ser, and Leu around the active site Ser; GRAVY, grand average of hydropathicity; Hconf, highly confident protein; HMA , heavy-metal-associated; HP, hypothetical protein; MCP , methyl-accepting chemotaxis protein; MgtE, magnesium transporter E; ORF , open reading frame; pdxH, pyridoxamine 5′-phosphate oxidase; pI, isoelectric point; PPI, protein–protein interaction; ROC, receiver operating characteristic; SAM-MT , S-adenosyl-L-methionine-dependent methyltransferase; TPR , tetratricopeptide repeat.

References

1 Kaakoush, N.O., Castaño-Rodríguez, N., Mitchell, H.M. and Man, S.M. (2015) Global epidemiology of Campylobacter infection. *Clin. Microbiol. Rev.* **28**, 687–720, https://doi.org/10.1128/CMR.00006-15

2 Platts-Mills, J.A. and Kosek, M. (2014) Update on the burden of Campylobacter in developing countries. *Curr. Opin. Infect. Dis.* **27**, 444, https://doi.org/10.1097/QCO.0000000000000091

3 Mehla, K. and Ramana, J. (2015) Novel drug targets for food-borne pathogen Campylobacter jejuni: an integrated subtractive genomics and comparative metabolic pathway study. *OMICS* **19**, 393–406, https://doi.org/10.1089/omi.2015.0046

4 Coker, A.O., Isokpehi, R.D., Thomas, B.N., Amisu, K.O. and Obi, C.L. (2002) Human campylobacteriosis in developing countries1. *Emerg. Infect. Dis.* **8**, 237, https://doi.org/10.3201/eid0803.010233

5 Takahashi, M., Koga, M., Yokoyama, K. and Yuki, N. (2005) Epidemiology of Campylobacter jejuni isolated from patients with Guillain-Barré and Fisher syndromes in Japan. *J. Clin. Microbiol.* **43**, 335–339, https://doi.org/10.1128/JCM.43.1.335-339.2005

6 Poly, F., Threadgill, D. and Stintzi, A. (2005) Genomic diversity in Campylobacter jejuni: identification of C. jejuni 81-176-specific genes. *J. Clin. Microbiol.* **43**, 2330–2338, https://doi.org/10.1128/JCM.43.5.2330-2338.2005

7 Parkhill, J., Wren, B., Mungall, K., Ketley, J., Churcher, C., Basham, D. et al. (2000) The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. *Nature* **403**, 665, https://doi.org/10.1038/35001088

8 Nimrod, G., Schushan, M., Steinberg, D.M. and Ben-Tal, N. (2008) Detection of functionally important regions in "hypothetical proteins" of known structure. *Structure* **16**, 1755–1763, https://doi.org/10.1016/j.str.2008.10.017

9 Shahbaaz, M., ImtaiyazHassan, M. and Ahmad, F. (2013) Functional annotation of conserved hypothetical proteins from Haemophilus influenzae Rd KW20. *PLoS ONE* **8**, e84263, https://doi.org/10.1371/journal.pone.0084263

10 Gazi, M.A., Kibria, M.G., Mahfuz, M., Islam, M.R., Ghosh, P., Afsar, M.N.A. et al. (2016) Functional, structural and epitopic prediction of hypothetical proteins of Mycobacterium tuberculosis H37Rv: an in silico approach for prioritizing the targets. *Gene* **591**, 442–455, https://doi.org/10.1016/j.gene.2016.06.057

11 da Costa, W.L.O., de Aragão Araújo, C.L., Dias, L.M., de Sousa Pereira, L.C., Alves, J.T.C., Araújo, F.A. et al. (2018) 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* **13**, e0198965, https://doi.org/10.1371/journal.pone.0198965

12 Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M. et al. (2016) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368, https://doi.org/10.1093/nar/gkw937

13 Snider, J., Kotlyar, M., Saraon, P., Yao, Z., Jurisica, I. and Stagljar, I. (2015) Fundamentals of protein interaction network mapping. *Mol. Syst. Biol.* **11**, 848, https://doi.org/10.15252/msb.20156351

14 Jez, J.M. (2017) Revisiting protein structure, function, and evolution in the genomic era. *J. Invertebr. Pathol.* **142**, 11–15, https://doi.org/10.1016/j.jip.2016.07.013

15 Gherardini, P.F. and Helmer-Citterich, M. (2008) Structure-based function prediction: approaches and applications. *Brief. Funct. Genom. Proteomics* **7**, 291–302, https://doi.org/10.1093/bfgp/eln030

16 Varma, P.B.S., Adimulam, Y.B. and Kodukula, S. (2015) In silico functional annotation of a hypothetical protein from Staphylococcus aureus. *J. Infect. Public Health* **8**, 526–532, https://doi.org/10.1016/j.jiph.2015.03.007

17 Ravooru, N., Ganji, S., Sathyanarayanan, N. and Nagendra, H.G. (2014) Insilico analysis of hypothetical proteins unveils putative metabolic pathways and essential genes in Leishmania donovani. *Front. Genet.* **5**, 291, https://doi.org/10.3389/fgene.2014.00291

18 Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y. et al. (2014) CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226, https://doi.org/10.1093/nar/gku1221

19 Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37, https://doi.org/10.1093/nar/gkr367

20 Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P. and Bork, P. (2000) SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**, 231–234, https://doi.org/10.1093/nar/28.1.231

21 Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R. et al. (2013) Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230, https://doi.org/10.1093/nar/gkt1223

22 De Castro, E., Sigrist, C.J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E. et al. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, W362–W365, https://doi.org/10.1093/nar/gkl124

23 Geer, L.Y., Domrachev, M., Lipman, D.J. and Bryant, S.H. (2002) CDART: protein homology by domain architecture. *Genome Res.* **12**, 1619–1623, https://doi.org/10.1101/gr.278202

24 Wilson, D., Madera, M., Vogel, C., Chothia, C. and Gough, J. (2006) The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Res.* **35**, D308–D313, https://doi.org/10.1093/nar/gkl910

25 Cai, C., Han, L., Ji, Z.L., Chen, X. and Chen, Y.Z. (2003) SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Res.* **31**, 3692–3697, https://doi.org/10.1093/nar/gkg600

26 Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L. et al. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208, https://doi.org/10.1093/nar/gkp335

27 Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J. et al. (2016) InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199, https://doi.org/10.1093/nar/gkw1107

28 Wei, W., Ning, L.-W., Ye, Y.-N. and Guo, F.-B. (2013) Geptop: a gene essentiality prediction tool for sequenced bacterial genomes based on orthology and phylogeny. *PLoS ONE* **8**, e72343, https://doi.org/10.1371/journal.pone.0072343

29 Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A. et al. (2010) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **39**, D1035–D1041, https://doi.org/10.1093/nar/gkq1126

30 Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, pp. 571–607, Springer, https://doi.org/10.1385/1-59259-890-0:571

31 Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R. et al. (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615, https://doi.org/10.1093/bioinformatics/btq249

32 Yu, C.S., Lin, C.J. and Hwang, J.K. (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* **13**, 1402–1406, https://doi.org/10.1110/ps.03479604

33 Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580, https://doi.org/10.1006/jmbi.2000.4315

34 Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378–379, https://doi.org/10.1093/bioinformatics/14.4.378

35 Tusnady, G.E. and Simon, I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–850, https://doi.org/10.1093/bioinformatics/17.9.849

36 Armenteros, J.J.A., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S. et al. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol* **37**, 420–423, https://doi.org/10.1038/s41587-019-0036-z

37 Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J. et al. (2014) STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452, https://doi.org/10.1093/nar/gku1003

38 Chen, C.-C., Hwang, J.-K. and Yang, J.-M. (2009) 2-v2: template-based protein structure prediction server. *BMC Bioinformatics* **10**, 366, https://doi.org/10.1186/1471-2105-10-366

39 Eng, J. (2017) *ROC Analysis: Web-based Calculator for ROC Curves*, Johns Hopkins University, Baltimore, MD, U.S.A., http://www.jrocfit.org

40 Pearson, W.R. (2013) An introduction to sequence similarity ("homology") searching. *Curr. Protoc. Bioinformatics* **42**, 3.1.8, https://doi.org/10.1002/0471250953.bi0301s42

41 Schnoes, A.M., Brown, S.D., Dodevski, I. and Babbitt, P.C. (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput. Biol.* **5**, e1000605, https://doi.org/10.1371/journal.pcbi.1000605

42 Gazi, M.A., Mahmud, S., Fahim, S.M., Kibria, M.G., Palit, P., Islam, M.R. et al. (2018) Functional prediction of hypothetical proteins from Shigella flexneri and validation of the predicted models by using ROC curve analysis. *Genomics Inform.* **16**, https://doi.org/10.5808/GI.2018.16.4.e26

43 Zhang, Z. and Ren, Q. (2015) Why are essential genes essential? The essentiality of Saccharomyces genes. *Microb. Cell* **2**, 280, https://doi.org/10.15698/mic2015.08.218

44 Romero, I.G., Ruvinsky, I. and Gilad, Y. (2012) Comparative studies of gene expression and the evolution of gene regulation. *Nat. Rev. Genet.* **13**, 505, https://doi.org/10.1038/nrg3229

45 Hajian-Tilaki, K. (2013) Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J. Int. Med.* **4**, 627

46 Verkhovsky, M.I. and Bogachev, A.V. (2010) Sodium-translocating NADH: quinone oxidoreductase as a redox-driven ion pump. *Biochim. Biophys. Acta* **1797**, 738–746, https://doi.org/10.1016/j.bbabio.2009.12.020

47 Wang, G. and Maier, R.J. (2004) An NADPH quinone reductase of Helicobacter pylori plays an important role in oxidative stress resistance and host colonization. *Infect. Immun.* **72**, 1391–1396, https://doi.org/10.1128/IAI.72.3.1391-1396.2004

48 Kroneck, P.M. and Torres, M.E.S. (2014) The metal-driven biogeochemistry of gaseous compounds in the environment. **14**, 333–335, https://doi.org/10.1007/978-94-017-9269-1

49 Simon, M.I., Borkovich, K.A., Bourret, R.B. and Hess, J.F. (1989) Protein phosphorylation in the bacterial chemotaxis system. *Biochimie* **71**, 1013–1019, https://doi.org/10.1016/0300-9084(89)90105-3

50 Hosfield, D.J., Mol, C.D., Shen, B. and Tainer, J.A. (1998) Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell* **95**, 135–146, https://doi.org/10.1016/S0092-8674(00)81789-4

51 S., K. (2006) Structure, function and mechanisms of action of ATPases from the AAA superfamily of proteins. *Postepy Biochem.* **52**, 330–338

52 Aravind, L., Leipe, D.D. and Koonin, E.V. (1998) Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.* **26**, 4205–4213, https://doi.org/10.1093/nar/26.18.4205

53 Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002) Introduction to pathogens. *Molecular Biology of the Cell*, 4th, Garland Science

54 Cerveny, L., Straskova, A., Dankova, V., Hartlova, A., Ceckova, M., Staud, F. et al. (2013) Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. *Infect. Immun.* **81**, 629–635, https://doi.org/10.1128/IAI.01035-12

55 Jönsson, K., Guo, B.P., Monstein, H.-J., Mekalanos, J.J. and Kronvall, G. (2004) Molecular cloning and characterization of two Helicobacter pylori genes coding for plasminogen-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1852–1857, https://doi.org/10.1073/pnas.0307329101

56 Zhang, J., Zhang, Y., Zhu, L., Suzuki, M. and Inouye, M. (2004) Interference of mRNA function by sequence-specific endoribonuclease PemK. *J. Biol. Chem.* **279**, 20678–20684, https://doi.org/10.1074/jbc.M314284200

57 Makhov, A., Hannah, J., Brennan, M., Trus, B., Kocsis, E., Conway, J. et al. (1994) Filamentous hemagglutinin of Bordetella pertussis: a bacterial adhesin formed as a 50-nm monomeric rigid rod based on a 19-residue repeat motif rich in beta strands and turns. *J. Mol. Biol.* **241**, 110–124, https://doi.org/10.1006/jmbi.1994.1478

58 Al-Khodor, S., Price, C.T., Kalia, A. and Kwaik, Y.A. (2010) Functional diversity of ankyrin repeats in microbial proteins. *Trends Microbiol.* **18**, 132–139, https://doi.org/10.1016/j.tim.2009.11.004

59 Kawai, F., Paek, S., Choi, K.-J., Prouty, M., Kanipes, M.I., Guerry, P. et al. (2012) Crystal structure of JlpA, a surface-exposed lipoprotein adhesin of Campylobacter jejuni. *J. Struct. Biol.* **177**, 583–588, https://doi.org/10.1016/j.jsb.2012.01.001

60 Tlapák, H., Rydzewski, K., Schulz, T., Weschka, D., Schunder, E. and Heuner, K. (2017) Functional analysis of the alternative sigma-28 factor FliA and its anti-sigma factor FlgM of the nonflagellated Legionella species L. oakridgensis. *J. Bacteriol.* **199**, e00018–17, https://doi.org/10.1128/JB.00018-17

61 Minamino, T., Chu, R., Yamaguchi, S. and Macnab, R.M. (2000) Role of FliJ in flagellar protein export in Salmonella. *J. Bacteriol.* **182**, 4207–4215, https://doi.org/10.1128/JB.182.15.4207-4215.2000

62 Christie, P.J. and Vogel, J.P. (2000) Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol.* **8**, 354–360, https://doi.org/10.1016/S0966-842X(00)01792-3

63 Natrajan, G., Noirot-Gros, M.F., Zawilak-Pawlik, A., Kapp, U. and Terradot, L. (2009) The structure of a DnaA/HobA complex from Helicobacter pylori provides insight into regulation of DNA replication in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21115–21120, https://doi.org/10.1073/pnas.0908966106

64 Roujeinikova, A. and Ud-Din AIMS (2017) Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. *Cell. Mol. Life Sci.* **74**, 3293–3303, https://doi.org/10.1007/s00018-017-2514-0

65 Yuan, J., Zweers, J.C., Van Dijl, J.M. and Dalbey, R.E. (2010) Protein transport across and into cell membranes in bacteria and archaea. *Cell. Mol. Life Sci.* **67**, 179–199, https://doi.org/10.1007/s00018-009-0160-x

66 Yan, Y.-W., Mao, D.-D., Yang, L., Qi, J.-L., Zhang, X.-X., Tang, Q.-L. et al. (2018) Magnesium transporter MGT6 plays an essential role in maintaining magnesium homeostasis and regulating high magnesium tolerance in Arabidopsis. *Front. Plant Sci.* **9**, 274, https://doi.org/10.3389/fpls.2018.00274

67 Bugde, P., Biswas, R., Merien, F., Lu, J., Liu, D.-X., Chen, M. et al. (2017) The therapeutic potential of targeting ABC transporters to combat multi-drug resistance. *Expert Opin. Ther. Targets* **21**, 511–530, https://doi.org/10.1080/14728222.2017.1310841

68 Bull, P.C. and Cox, D.W. (1994) Wilson disease and Menkes disease: new handles on heavy-metal transport. *Trends Genet.* **10**, 246–252, https://doi.org/10.1016/0168-9525(94)90172-4

69 Bailey, S., Ward, D., Middleton, R., Grossmann, J.G. and Zambryski, P.C. (2006) Agrobacterium tumefaciens VirB8 structure reveals potential protein–protein interaction sites. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2582–2587, https://doi.org/10.1073/pnas.0511216103

70 Pierson, III, L.S., Gaffney, T., Lam, S. and Gong, F. (1995) Molecular analysis of genes encoding phenazine biosynthesis in the biological control bacterium Pseudomonas aureofaciens 30-84. *FEMS Microbiol. Lett.* **134**, 299–307

71 Hunter, R.C., Klepac-Ceraj, V., Lorenzi, M.M., Grotzinger, H., Martin, T.R. and Newman, D.K. (2012) Phenazine content in the cystic fibrosis respiratory tract negatively correlates with lung function and microbial complexity. *Am. J. Respir. Cell Mol. Biol.* **47**, 738–745, https://doi.org/10.1165/rcmb.2012-0088OC

72 Upadhyay, A. and Srivastava, S. (2011) Phenazine-1-carboxylic acid is a more important contributor to biocontrol Fusarium oxysporum than pyrrolnitrin in Pseudomonas fluorescens strain Psd. *Microbiol. Res.* **166**, 323–335, https://doi.org/10.1016/j.micres.2010.06.001

73 Banco, M.T., Mishra, V., Greeley, S.C. and Ronning, D.R. (2018) Direct detection of products from S-adenosylmethionine-dependent enzymes using a competitive fluorescence polarization assay. *Anal. Chem.* **90**, 1740–1747, https://doi.org/10.1021/acs.analchem.7b03556

74 Martin, J.L. and McMillan, F.M. (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **12**, 783–793, https://doi.org/10.1016/S0959-440X(02)00391-3

75 Ruan, J., Wu, X., Ye, Y. and Härdter, R. (1998) Effect of potassium, magnesium and sulphur applied in different forms of fertilisers on free amino acid content in leaves of tea (Camellia sinensis L). *J. Sci. Food Agric.* **76**, 389–396, https://doi.org/10.1002/(SICI)1097-0010(199803)76:3%3c389::AID-JSFA963%3e3.0.CO;2-X

76 Frias, J., Doblado, R., Antezana, J.R. and Vidal-Valverde, C. (2003) Inositol phosphate degradation by the action of phytase enzyme in legume seeds. *Food Chem.* **81**, 233–239, https://doi.org/10.1016/S0308-8146(02)00417-X

77 Dersjant-Li, Y., Awati, A., Schulze, H. and Partridge, G. (2015) Phytase in non-ruminant animal nutrition: a critical review on phytase activities in the gastrointestinal tract and influencing factors. *J. Sci. Food Agric.* **95**, 878–896, https://doi.org/10.1002/jsfa.6998

78 Harada, J., Takaku, S. and Watanabe, K. (2012) An on-demand metalloprotease from psychro-tolerant Exiguobacterium undae Su-1, the activity and stability of which are controlled by the Ca2+ concentration. *Biosci. Biotechnol. Biochem.* **76**, 986–992, https://doi.org/10.1271/bbb.110997

79 Akoh, C.C., Lee, G.-C., Liaw, Y.-C., Huang, T.-H. and Shaw, J.-F. (2004) GDSL family of serine esterases/lipases. *Prog. Lipid Res.* **43**, 534–552, https://doi.org/10.1016/j.plipres.2004.09.002

80 KONG, K.F., Schneper, L. and Mathee, K. (2010) Beta-lactam antibiotics: from antibiosis to resistance and bacteriology. *APMIS* **118**, 1–36, https://doi.org/10.1111/j.1600-0463.2009.02563.x

81 Lam, W.W.L., Woo, E.J., Kotaka, M., Tam, W.K., Leung, Y.C., Ling, T.K.W. et al. (2010) Molecular interaction of flagellar export chaperone FliS and cochaperone HP1076 in Helicobacter pylori. *FASEB J.* **24**, 4020–4032, https://doi.org/10.1096/fj.10-155242

82 Meganathan, R. (2001) Biosynthesis of menaquinone (vitamin K2) and ubiquinone (coenzyme Q): a perspective on enzymatic mechanisms. *Vitamins Hormones* **61**, 173–218

83 Debnath, J., Siricilla, S., Wan, B., Crick, D.C., Lenaerts, A.J., Franzblau, S.G. et al. (2012) Discovery of selective menaquinone biosynthesis inhibitors against Mycobacterium tuberculosis. *J. Med. Chem.* **55**, 3739–3755, https://doi.org/10.1021/jm201608g

84 Idicula-Thomas, S. and Balaji, P.V. (2005) Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli. *Protein Sci.* **14**, 582–592, https://doi.org/10.1110/ps.041009005

85 Jaspard, E., Macherel, D. and Hunault, G. (2012) Computational and statistical analyses of amino acid usage and physico-chemical properties of the twelve late embryogenesis abundant protein classes. *PLoS ONE* **7**, e36968, https://doi.org/10.1371/journal.pone.0036968

86 Meier, M., Sit, R.V. and Quake, S.R. (2013) Proteome-wide protein interaction measurements of bacterial proteins of unknown function. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 477–482, https://doi.org/10.1073/pnas.1210634110

87 Chance, M.R., Bresnick, A.R., Burley, S.K., Jiang, J.S., Lima, C.D., Sali, A. et al. (2002) Structural genomics: a pipeline for providing structures for the biologist. *Protein Sci.* **11**, 723–738, https://doi.org/10.1110/ps.4570102

88 Cardew, E.M., Verlinde, C.L. and Pohl, E. (2018) The calcium-dependent protein kinase 1 from Toxoplasma gondii as target for structure-based drug design. *Parasitology* **145**, 210–218, https://doi.org/10.1017/S0031182017001901

89 Kumar, A., Sharma, A., Kaur, G., Makkar, P. and Kaur, J. (2017) Functional characterization of hypothetical proteins of Mycobacterium tuberculosis with possible esterase/lipase signature: a cumulative in silico and in vitro approach. *J. Biomol. Struct. Dyn.* **35**, 1226–1243, https://doi.org/10.1080/07391102.2016.1174738

90 Choi, H.-P., Juarez, S., Ciordia, S., Fernandez, M., Bargiela, R., Albar, J.P. et al. (2013) Biochemical characterization of hypothetical proteins from Helicobacter pylori. *PLoS ONE.* **8**, https://doi.org/10.1371/journal.pone.0066605

91 Cort, J.R., Yee, A., Edwards, A.M., Arrowsmith, C.H. and Kennedy, M.A. (2000) NMR structure determination and structure-based functional characterization of conserved hypothetical protein MTH1175 from Methanobacterium thermoautotrophicum. *J. Struct. Funct. Genomics* **1**, 15–25, https://doi.org/10.1023/A:1011348803324

92 Barta, M.L., Thomas, K., Yuan, H., Lovell, S., Battaile, K.P., Schramm, V.L. et al. (2014) Structural and biochemical characterization of Chlamydia trachomatis hypothetical protein CT263 supports that menaquinone synthesis occurs through the futalosine pathway. *J. Biol. Chem.* **289**, 32214–32229, https://doi.org/10.1074/jbc.M114.594325

93 Zhang, W., Culley, D.E., Gritsenko, M.A., Moore, R.J., Nie, L., Scholten, J.C. et al. (2006) LC–MS/MS based proteomic analysis and functional inference of hypothetical proteins in Desulfovibrio vulgaris. *Biochem. Biophys. Res. Commun.* **349**, 1412–1419, https://doi.org/10.1016/j.bbrc.2006.09.019