Pathway based therapeutic targets identification and development of an interactive database CampyNIBase of *Campylobacter jejuni* RM1221 through non-redundant protein dataset

Mohammad Uzzal Hossain[1], Taimur Md. Omar[2], Iftekhar Alam[3], Keshob Chandra Das[4], A. K. M. Mohiuddin[2], Chaman Ara Keya[5], Md. Salimullah[4]*

1 Bioinformatics Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka, Bangladesh, **2** Department of Biotechnology and Genetic Engineering, Life Science Faculty, Mawlana Bhashani Science and Technology University, Santosh, Tangail, Bangladesh, **3** Plant Biotechnology Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka, Bangladesh, **4** Molecular Biotechnology Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka, Bangladesh, **5** Department of Biochemistry and Microbiology, North south University, Bashundhara, Dhaka, Bangladesh

* salim2969@gmail.com

Abstract

The bacterial species *Campylobacter jejuni RM1221* (*CjR*) is the primary cause of campylobacteriosis which poses a global threat for human health. Over the years the efficacy of antibiotic treatment is becoming more fruitless due to the development of multiple drug resistant strains. Therefore, identification of new drug targets is a valuable tool for the development of new treatments for affected patients and can be obtained by targeting essential protein(s) of *CjR*. We conducted this *in silico* study in order to identify therapeutic targets by subtractive *CjR* proteome analysis. The most important proteins of the *CjR* proteome, which includes chokepoint enzymes, plasmid, virulence and antibiotic resistant proteins were annotated and subjected to subtractive analyses to filter out the *CjR* essential proteins from duplicate or human homologous proteins. Through the subtractive and characterization analysis we have identified 38 eligible therapeutic targets including 1 potential vaccine target. Also, 12 potential targets were found in interactive network, 5 targets to be dealt with FDA approved drugs and one pathway as potential pathway based drug target. In addition, a comprehensive database 'CampyNIBase' has also been developed. Besides the results of this study, the database is enriched with other information such as 3D models of the identified targets, experimental structures and Expressed Sequence Tag (EST) sequences. This study, including the database might be exploited for future research and the identification of effective therapeutics against campylobacteriosis. URL: (http://nib.portal.gov.bd/site/page/4516e965-8935-4129-8c3f-df95e754c562#Banner).

Introduction

The genus *Campylobacter* is composed of a wide variety of non-spore forming Gram-negative bacteria that are predominantly rod or spiral shaped [1]. Most *Campylobacter* infections are acquired through contaminated food and two species, *Campylobacter coli* and *Campylobacter jejuni* are the primary cause of the human disease termed as campylobacteriosis [2,3]. *Campylobacter jejuni (C. jejuni)* is the species that induces acute gastroenteritis and bacterial food poisoning in infected patients. Normal infection with *C. jejuni* causes uncomplicated gastroenteritis but severe infection may result in abdominal cramps, fever or even serious diseases like diarrhea, GuillainBarré Syndrome or Miller Fischer Syndrome [4–6]. *C. jejuni* infection is acquired via numerous sources associated with lack of awareness such as undercooked livestock meat, poultry, unpasteurized milk or contaminated water sources [7].

C. jejuni causes the highest proportion of campylobacteriosis cases in developed countries, and in the United States, between $1.3 to $6.8 billion dollars is spent annually for treating the illness [8,9]. According to the ECDC (European Centre for Disease Prevention and Control) and EFSA (European Food Safety Authorities) report the most important zoonosis was found to be campylobacteriosis compared to yersinosis and salmonellosis [10–12]. Global campylobacteriosis incidence is increasing each year and has almost exceeded the incidence of shigella infections [13,14]. In 2010 New Zealand endured one of the highest rates of campylobacteriosis demonstrating that *Campylobacter* infection is a global threat [10].

CjR is a disease causing strain which has similar *C. jejuni* type of infectious properties. Antibiotic treatment against *CjR* is becoming increasingly more ineffective due to the emergence of multiple antibiotic resistance strains. This resistance requires special attention as *C. jejuni* is capable of efficient transfer of the resistant genes into other strains. In 2000, data of primary genome sequence of *C. jejuni* was released but detailed information about variations and polymorphisms in the complete genome sequences of different strains was only published in 2006 [15–17]. Currently, information about gene/protein sequences and metabolic pathways of *CjR* is available in various databases like NCBI, KEGG, Biocyc.org etc. These databases have become a critical tool for the discovery and identification of new molecular target(s) and subsequently provide a valuable platform for researchers and the pharmaceutical industry to enable the development of new drugs and vaccines. Targets should normally be an eligible gene or protein of a specific strain which can be targetable by an existing or non-existing drug.

Conventional drug target identification is time consuming, expensive, laborious and often only a few drug targets can be identified. In comparison, the *in silico* approach allows for a great deal of analysis to be carried out within a short period of time which is cost effective and often delivers a high number of the promising drug targets from a large pool of information. This has been facilitated by information available from various databases that provide whole genome sequences of various organisms ranging from pathogenic bacteria to human [18,19]. Currently, utilization of various *in silico* approaches to identify potential vaccine or drug targets has become a prerequisite for drug and vaccine design [20–22]. One such *in silico* approach is Subtractive genomic analysis which attempts to discover new proteins or targets that are important for the survival of the pathogenic microorganism and non-homologous to the human host. Designing drugs against these proteins have a high probability to be effective against their target microorganism [23]. In this study, we applied this subtractive proteome analysis to identify essential proteins such as chokepoint enzymes, virulent proteins and antibiotic resistant proteins that are important for the survival of *CjR*. This Subtractive analysis is possible due to the genome of *CjR* becoming available in various web based databanks. After identification and confirmation of human non-homologues drug targets are characterized to facilitate effective drug design. Additionally, we have also analyzed and suggested pathways for

future drug targets and created an open access database named CampyNIBase to store all identified and characterized therapeutic targets. Furthermore, in order to enrich the database, other relevant campylobacter information such as expressed sequence tags (EST) sequences, experimental drugs are incorporated in CampyNIBase through extensive surveying of the literature.

Materials and methods

The identification of drug targets was carried out in three phases. In phase-I, major proteins of *CjR* genome were collected from different sources. These proteins include chokepoint enzymes, plasmid proteins, virulent proteins and antibiotic resistant proteins. In phase-II, subtractive analyses were carried out through different steps by excluding human homologues proteins and collecting important proteins required for the survival of *CjR*. In phase-III, potential drug targets found from subtractive analyses in phase-II are characterized. The entire work flow can be seen as a flowchart (Fig 1).

Phase–I: Collection of *CjR* proteins

Analyzing chokepoint enzymes.   Chokepoint enzymes are available in Biocyc (Biocyc. org) [24], a server containing a collection of Pathway/Genome Databases (PGDBs) of various organisms. In the database server, *CjR* was selected as desired. Then each of the enzymes involved in chokepoint reactions on the consuming side and producing side was collected. In this stage, only EC numbers and names of the enzymes were collected. By the same process, those of human chokepoint enzymes were also collected. This is a vital step as proteins with similar domains to a human enzyme may lead to harmful drug or vaccine side effects in treated patients. Subsequently enzymes of *CjR* similar to proteins expressed by humans were excluded to avoid cross targeting. Next the EC numbers or names collected more than once were rejected to reduce manual work. Finally, FASTA sequences were obtained and only the enzymes unique to the *CjR* strain were collected. A list of the selected enzymes was collected as (α3) list (Fig 2).

Plasmid protein retrieval.   Plasmids of *C. jejuni* can contain unique characteristics compared to other *Campylobacter* species. Plasmid sequences were retrieved as those characteristics can also be potential drug targets. The protein sequences were retrieved from literature review [25]. Only the unique proteins of *CjR* were collected and listed as delta (δ) list (Fig 2).

Analyzing virulence factor.   Virulence factors expressed by bacteria are required for the survival and success of pathogenic bacteria and are potential drug targets. They were identified and collected from the virulence factor database (VFDB) [26]. To collect all known virulence factors of *CjR*, the NCBI database [27] linked with VFDB was useful. All the retrieved proteins were listed as beta (β) list: Proteins unique to *CjR1* strain were selected in (β1) list and other proteins were excluded and listed in (β1') list (Fig 2).

Analyzing antibiotic resistant genes.   Many of the strains of *C. jejuni* have developed resistance to multiple drugs. Genetic mutations in certain genes of *C. jejuni* are responsible for such kind of resistance [28–30]. The proteins interacting with these resistant gene components were identified by web based tool, Search Tool for the Retrieval of Interacting Genes (STRING 10) [31]. Common proteins found in all *C. jejuni* strains that were available in the STRING 10 database and known for interacting with resistant genes were collected in Excel file. The FASTA sequences of these proteins were retrieved from NCBI protein database [27] and some additional resistant proteins were retrieved from Antibiotic Resistance Genes Database (ARDB) [32]. The FASTA sequence of all of the proteins was collected and listed as gamma (γ) list (Fig 2). Only the proteins unique to *CjR* were collected and others were avoided (Fig 2).
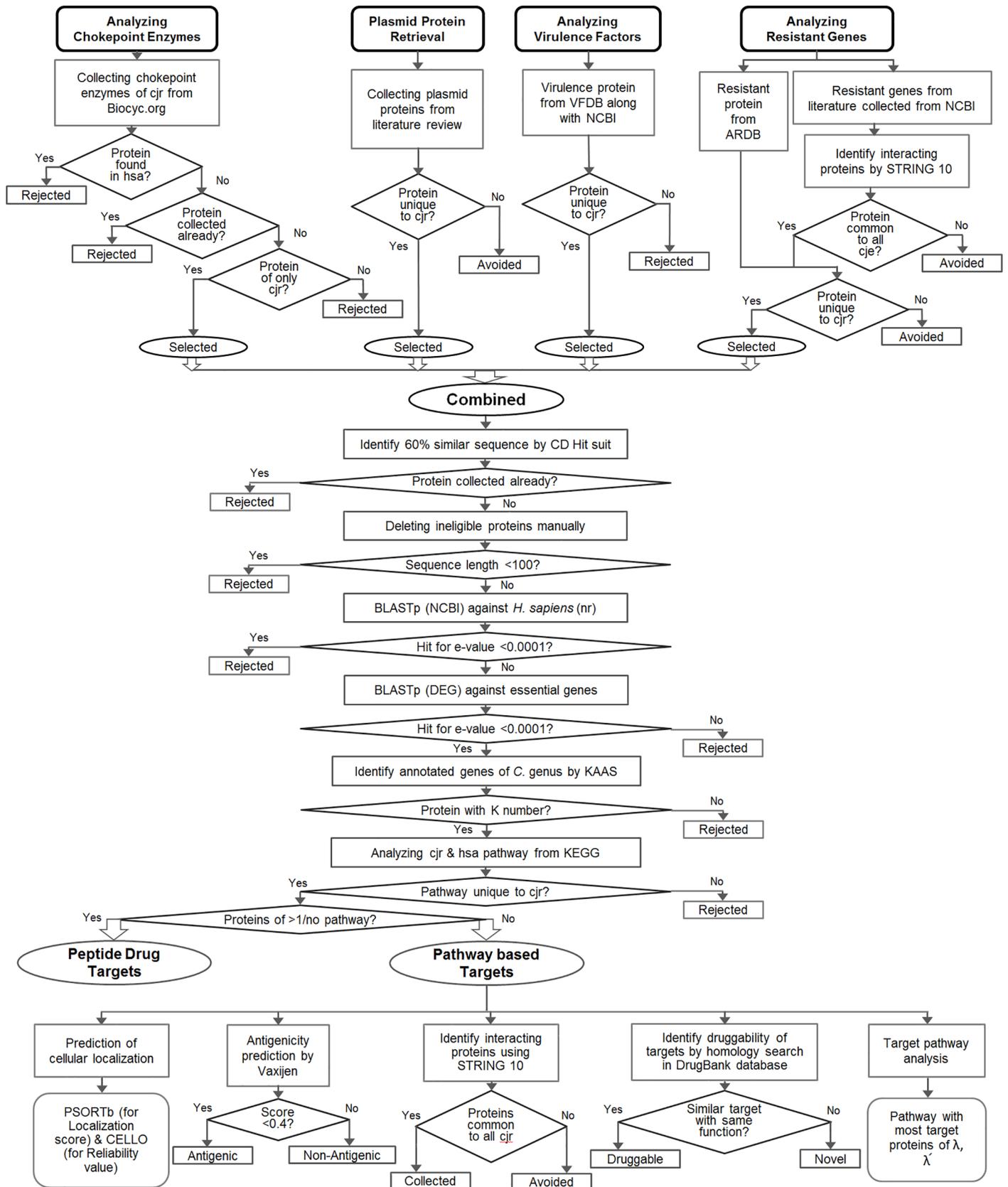
Phase–II: Subtractive analysis of drug targets

Duplicate sequences removal. As the FASTA sequences of the proteins were collected from different sources, many of these were collected more than once. Thus a sequence clustering web server, CD-HIT Suite [33] was used to compare all the retrieved sequences and remove duplicates. The input was 60% in sequence identity as the cut-off parameter and selected proteins were collected in (π) list (Fig 2).
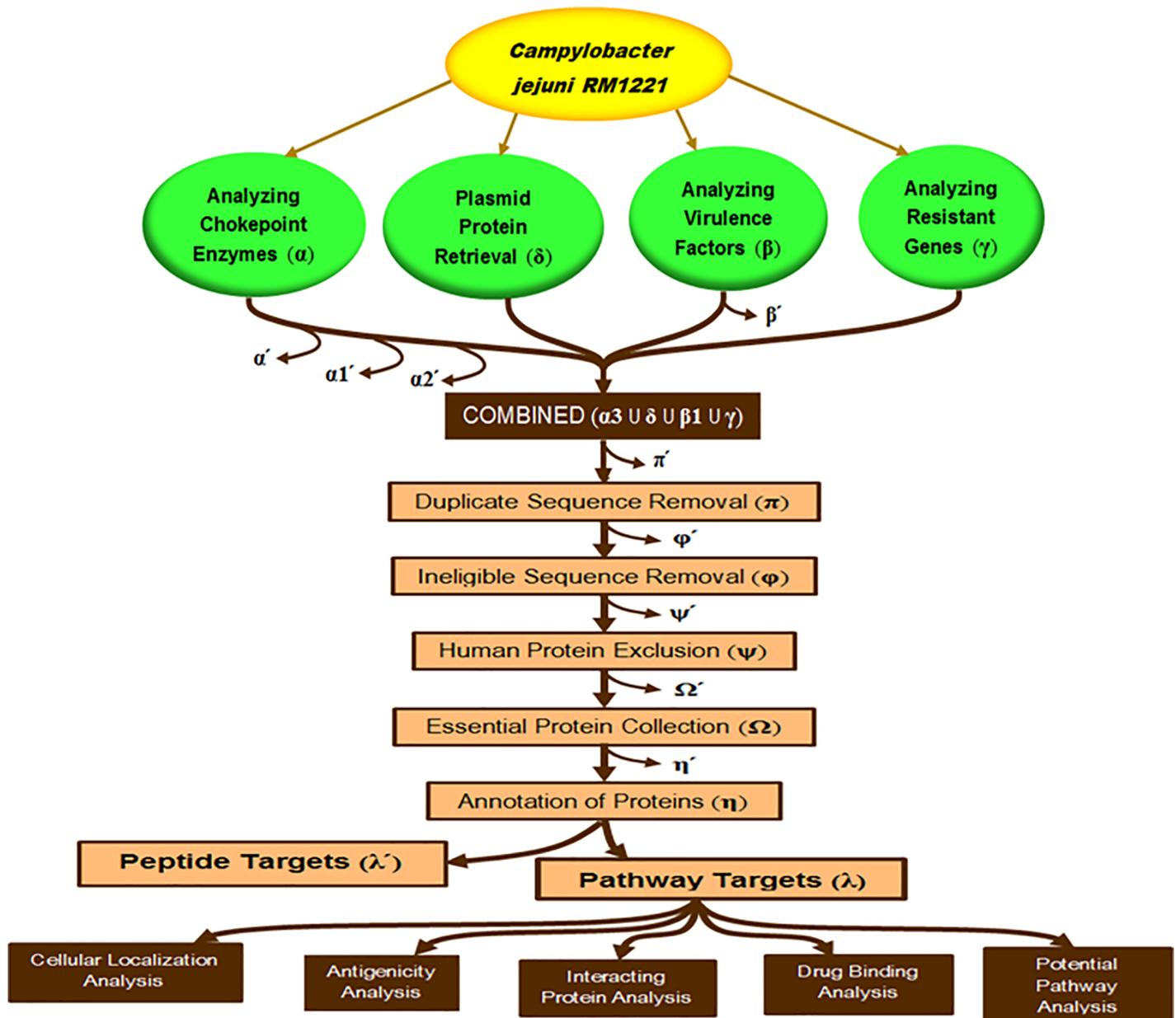
Fig 2. Outline of subtractive analysis and list of proteins. The enlisted proteins were prioritized (different symbol) from different phase of subtractive channel of analysis.

https://doi.org/10.1371/journal.pone.0198170.g002

Ineligible sequence removal. Proteins with sequence length less than 100 were considered ineligible, removed manually and listed in (φ´) list, whereas proteins with length ≥ 100 were enlisted in (φ) list (Fig 2).

Human protein exclusion. If the target protein of *CjR* is homologous to a human protein, then the drug or vaccine developed against it has potential to also target the human protein. Hence, proteins homologous to human proteins were removed. For this purpose, proteins of (φ) list were submitted to BLASTp [34] against a non-redundant database of *Homo sapiens*. The considered threshold was 10^{-3}. The proteins non-homologous to human were collected and listed as (ψ) list (Fig 2).

Essential protein collection. *C. jejuni* is one of the leading causes of food poisoning and diarrhea [4–6]. Therefore, other strains responsible for food poisoning and diarrhea were selected for essential protein analysis. The major bacteria responsible for food poisoning are *Clostridium perfringens*[35], *Salmonella spp*. [36], *Escherichia coliO157:H7* [37], *Bacillus cereus* [38], *Listeria monocytogenes*[39], *Shigella spp*. [40], *Staphylococcus aureus*[41], *Staphylococcal enteritis*[42], *Streptococcus*[43,44], *Vibrio cholerae*[45], *Vibrio parahaemolyticus*[46], *Vibrio vulnificus*[47], *Yersinia* enterocolitica [48] *Brucella spp*[49], *Coxiella*burnetii [50] and *Plesiomonas* shigelloides [51]. *Escherichia coli*, *Salmonella*, *Shigella*, *Clostridium* difficile [52], *Staphylococcus aureus*[53] can cause diarrhea. A BLAST search was performed for proteins of (ψ) list against the essential proteins of common organisms mentioned above from the DEG 10 [54] database. BLASTp program analysis was performed in Protein Query vs. Protein Database. The expect value was taken as 1E-10 and essential proteins were enlisted in (Ω) list (Fig 2).

Annotation of proteins. In the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [55], genes in complete genomes are annotated with the KEGG orthology (KO) identifiers, or the K numbers. The essence of the KO system is that it is a pathway based definition of orthologous genes. The KO entry represents an ortholog group that is linked to a box (gene product) in the KEGG pathway diagram. Thus, once the KO identifiers, or the K numbers are assigned to genes in the genome, which is manually verified in KEGG, organism-specific pathways can be computationally generated. Proteins of (Ω) list were subjected to BLASTp against 10 strains of "Campylobacter" found in KEGG Automatic Annotation Server (KAAS) [56]. Proteins with KO number were collected and others were excluded. Selected proteins were enlisted in (η) list with KO numbers (Fig 2).

Pathway analysis. Pathways of the *CjR* proteins were revealed from the KEGG [55] database. If a protein of *C. jejuni*is involved in the same pathway found in humans then the drug target can be harmful for the host. Therefore proteins common to *C. jejuni* and *H. sapiens* pathway were excluded to avoid cross targeting and only proteins unique to pathways in *C. jejuni* were collected. To find out pathway based drug targets, proteins involved in only one pathway were collected and enlisted in (λ) list (Fig 2).

Phase–III: Characterization of pathway based drug targets

A pathway based drug is a drug which targets whole pathway rather than single protein. Proteins known to be involved in one pathway were considered as major targets as targeting one pathway is easier to produce a pathway based drug. These are the proteins enlisted in lambda (λ) list and several characteristic features of these proteins were figured out.

Subcellular location analysis. Proteins expressed on the surface of bacteria are probable vaccine antigen candidates while proteins found in the inner membrane regions and cytoplasm of the bacteria are probable drug candidates. Thus, knowing the locations of the proteins is extremely useful and necessary. A bacterial protein subcellular localization (SCL) predictor,

PSORTb 3.0 [57], and CELLO v.2.5 [58], a server based on a two-level support vector machine (SVM) system predicting subcellular localization of the proteins were used. As a protein's function is related to its localization, this prediction would be used for functional analysis.

Antigenicity analysis. To predict protective antigens of bacteria, Vaxijen [59], an alignment-independent prediction server, was used. It reveals the antigenic score of every protein. The proteins with higher antigenic score were considered as a more viable antigen target. Proteins with antigenic score less than 0.4 (default threshold value) were considered as non-antigenic proteins. We used NetCTL 1.2 server (http://www.cbs.dtu.dk/services/NetCTL/) and AllerTOP v.2.0 server (http://www.ddg-pharmfac.net/AllerTOP/) for the identification of potential T- cell epitopes from the proteins that had high antigenic scores. Immune Epitope Database (IEDB) [60] tools were utilized for MHC-I molecules interactions with potential T-cell epitopes as well as epitope conservancy analysis. Further, we employed a set of bioinformatics tools including Kolaskar and Tongaonkar antigenicity scale [61], Emini surface accessibility prediction [62], Karplus and Schulz flexibility prediction [63], Bepipred linear epitope prediction analysis [64] and Chou and Fasman beta turn prediction analysis [65, 66] to predict the B-cell antigenicity.

Interacting protein analysis. A protein functions in a system of an organism by interacting with other proteins. To characterize such interacting proteins of each target of (λ) list STRING 10 [31], a web based search tool, was used. As this tool does not have information about *CjR* strain, we collected common interacting genes among available *C. jejuni* strains for each target. The target with a greater number of edges, interacts with more genes (nodes) and has a higher importance in a system. If a target forms numerous interactions with other proteins, it might be termed as 'hub' [31].

Drug binding analysis. Another BLASTp program was setup to know whether targets can be treated by current Food and Drug Administration (FDA) approved drugs. So, each target in λ list were subjected to BLASTp against DrugBank 3.0 target collection [67] for homology comparisons. The hit targets from λ list with DrugBank can be dealt with approved and available drugs.These targets were named as 'Druggable' targets. Targets nonhomologous to targets of the DrugBank 3.0 target collection were termed as 'Novel' targets. Furtheremore, Autodock Vina [68] was utilized to perform the blind docking to predict the binding affinity between druggable targets and predicted drug molecules.

Potential pathway analysis. Drugs can be designed to target whole pathways rather than a single protein. So, a pathway containing more targets of η list was considered as a pathway with greater potential. Here, we only took pathways of λ list for manual analysis. Pathways containing proteins involved in human pathways were ignored. Pathways containing the highest number of proteins of η list were considered as the most potential pathway.

Experimental and tertiary structure identification. The PDB [69] database was scanned to identify the experimental structures with the query of identified sequences of therapeutic targets. The available structures were deposited in our database 'CampyNIBase' whilst the remaining sequences that showed no hit to the PDB database were also employed to build 3D structures for facilitating the drug discovery. We have selected the best template from the Local meta-threading-server (LOMETS) [70] where more than one threading program showed the same template. Modeller 9.17 [71] was used for generating a number of models. Subsequently, DOPE (Discrete Optimized Protein Energy) scoring was considered to select the best model from a number of generated models. Thereafter, quality assessment of built models was checked by Ramachandran Plots [72]. We also employed the COFACTOR server [73] for the prediction of binding site in the generated models.

Database development and organization. We have developed a user friendly open access database named CampyNIBase in which all the identified drug targets were deposited. The

storage system for this database was based on My-SQL hosted by Bangladesh Computer Council Data Centre. The user interface and back-end of this database was based on an open source scripting language PHP. The contents of our developed database were categorized with different types of menu bar.

Results

Annotation of *CjR* proteome

Data retrieved from the Biocyc database revealed that *CjR* contains 238 and 251 chokepoint reactions on the consuming and producing side respectively [24]. Among them, 236 chokepoint enzymes were involved in consuming chokepoint reactions and 260 chokepoint enzymes were involved in producing chokepoint reactions. Thus, the number of total consuming and producing enzymes is 496 collected in alpha (α) list. After excluding 195 human enzymes (α′ list), 301 chokepoint enzymes remained (α1 list). 107 enzymes (α1′ list) were found to be collected more than once and after their removal, 194 enzymes (α2 list) remained. While collecting the FASTA sequences, it was found that all of these enzymes were not unique to the *CjR* strain. Therefore following the exclusion of 91 proteins (α2′ list) of other strains, 103 chokepoint enzymes (α3 list) were selected to remain (Table 1, S1 Table). Two plasmid proteins of *Campylobacter* were found from literature review [25]. Among them one was specific for *CjR*. This protein was enlisted as delta (δ) list (Table 1, S2 Table). A total of 2204 virulence proteins

Table 1. List of targets through identification channel.

Phase	Analysis	Process	Collected for Analysis	Rejected	Selected	Final selected proteins from a phase
I	Analyzing chokepoint enzymes	(i) Rejecting proteins of hsa	**496(α)**	**195(α′)**	**301(α1)**	**2218**
		(ii) Rejecting duplicate proteins	**301(α1)**	**107(α1′)**	**194(α2)**	
		(iii) Selecting unique proteins of cjr	**194(α2)**	**91(α2′)**	**103(α3)**	
	Plasmid protein retrieval	Collecting protein from literature review	**1(δ)**	**0**	**1(δ)**	
	Analyzing virulence factor	Selecting unique proteins of cjr	**2204(β)**	**104(β′)**	**2100(β1)**	
	Analyzing antibiotic resistant genes	Selecting unique proteins of cjr	**14(γ)**	**0**	**14(γ)**	
II	Duplicate sequence removal	60% sequence identity cut-off	**2218**	**407(π′)**	**1811(π)**	**38**
	Ineligible protein removal	Removing sequence length <100	**1811**	**240(φ′)**	**1571(φ)**	
	Human protein exclusion	Removing proteins homologous to Human by BLASTp	**1571**	**384(ψ′)**	**1187(ψ)**	
	Essential protein collection	Collecting essential proteins by BLASTp against DEG database	**1187**	**802(Ω′)**	**385(Ω)**	
	Annotation of proteins	Proteins with KO number collection from KAAS	**385**	**67(η′)**	**318(η)**	
	Pathway analysis	Identify unique pathway and proteins of 1 pathway from KEGG	**318**	**280(λ′)**	**38(λ)**	
III	Subcellular Location Analysis	PSORTb and CELLO protein location	**38(λ)**			
	Antigenicity Analysis	Vaxijen score for antigenicity	**38(λ)**			
	Broad Spectrum Analysis	Interacting protein analysis	**38(λ)**			
	Druggability Analysis	Targets in Drug Bank	**38(λ)**			

were collected from VFDB [26] along with NCBI database [27] and assigned to beta (β) list. Among them 104 proteins were manually identified as common to other strains of *C. jejuni* and collected in (β′) list followed by rejection. Therefore, 2100 proteins unique to *CjR* were collected and enlisted in (β1) list (Table 1, S3 Table). From literature reviews, the cmeB [28], gyrA [29] and *aphA*-7 [30] antibiotic resistant genes were collected. In *C. jejuni*, *cmeB* and *gyrA* were found to interact directly with 1 and 5 proteins of *CjR* respectively but *aphA-7* did not interact with other proteins. In this step, a total of 9 proteins were found as the number of resistant proteins along with interacting proteins. Also from VFDB [26], 5 proteins were retrieved. So, the number of total proteins from this analysis is 14 and they were listed in gamma (γ) list (Table 1, S4 Table). All of the chokepoint enzymes, plasmid proteins, virulent proteins and antibiotic resistant proteins listed in α3, δ, β1 and γ lists respectively were combined (S5 Table) to obtain a total of 2218 proteins for subtractive analyses.

Identification of potential therapeutic targets

Recognition of human non-homologous protein of *CjR*.   At the beginning of therapeutic target identification, we have removed 407 duplicate protein sequences from the 2218 proteins obtained during the annotation of the *CjR* proteins. The duplicate sequences are listed in (π′) list. The remaining 1811 sequences listed in (π) list contains 101 chokepoint enzymes, 1 plasmid protein, 1704 virulent proteins and 5 antibiotic resistant proteins (Table 1, S6 Table). Next we manually excluded 240 (φ′) proteins with length <100 amino acids from (π) list and 1571 proteins were collected and listed in (φ) list. There are 101 chokepoint enzymes, 1465 virulent proteins and 5 antibiotic resistant proteins in (φ) list (Table 1, S7 Table). Finally, through BLASTp program [34], 1187 proteins were found to be non-homologous to proteins expressed by humans. Those proteins were retrieved and and listed (ψ). The ψ list contains 75 chokepoint,1108 virulent and 4 antibiotic resistant proteins. The remaining 384 proteins were excluded and listed as (ψ′) list (Table 1, S8 Table).

CjR specific essential proteins.   Essential proteins are proteins that are required for the survival of a species in any condition. Therefore, we attempted to identify *CjR* specific essential proteins. Through DEG 10 database [54] 16 bacterial species were found to cause food poisoning and diarrhea. They are *Bacillus subtilis 168*, *Campylobacter jejuni NCTC 11168*, *Escherichia coli MG1655 I*, *Escherichia coli MG1655 II*, *Salmonella entericaserovar Typhi*, *Pseudomonas aeruginosa PAO1*, *Salmonella entericaserovar Typhimurium SL1344*, *Salmonella entericaserovar Typhimurium 14028S*, *Salmonella typhimurium LT2*, *Staphylococcus aureus N315*, *Staphylococcus aureus NCTC 8325*, *Streptococcus pneumonia*, *Streptococcus pyogenes MGAS5448*, *Streptococcus pyogenes NZ131*, *Streptococcus sanguinis* and *Vibrio cholerae N16961*. After protein BLAST against DEG 10 database 385 proteins were found as essential proteins for CjR and contained 45 chokepoint proteins, 338 virulent factors and 2 antibiotic resistant proteins which are listed in (Ω) list. The remaining 802 proteins were rejected and listed in (Ω′) list (Table 1, S9 Table).

Pathway based identification of therapeutic targets.   385 *CjR* specific essential proteins were processed with KEGG Automatic Annotated Server (KAAS). Following the processing procedure, 318 proteins were found in KEGG Ortholog (KO) list with K numbers and collected in (η) list containing 45 chokepoint proteins, 271 virulence proteins and 2 antibiotic resistant proteins. The other 67 proteins were removed and listed as (η′) list (Table 1, S10 Table).Using the KEGG[55] database, 26 proteins of (η) list were found to be involved in human pathways and were removed. 38 proteins were found in only one pathway and selected as potential therapeutic targets (λ) list. Other proteins were enlisted in (λ′) list (Table 2, S11 and S12 Tables).

Table 2. List of pathway based drug targets.

| No.	KEGG ID	Protein Name	KO Number	Pathway	Pathway ID	NCBI Protein ID	Length (a.a.)	Chokepoint Enzyme	Virulence Factor
1	CJE1596	holo-(acyl-carrier-protein) synthase	K00997	Pantothenate and CoA biosynthesis	cjr00770	AAW36029	115	Yes	No
2	CJE1405	N-acetylmuramoyl-L-alanine amidase	K01448	Cationic antimicrobial peptide resistance	cjr01503	AAW35724	659	Yes	No
3	CJE1810	DNA primase	K02316	DNA replication	cjr03030	AAW36232	605	No	Yes
4	CJE1654	ATP-dependent DNA helicase, UvrD/REP family	K03582	Homologous recombination	cjr03440	AAW36087	921	No	Yes
5	CJE1566	non-canonical purine NTP pyrophosphatase, RdgB/HAM1 family	K02428	Purine metabolism	cjr00230	AAW36001	200	No	Yes
6	CJE1554	Holliday junction DNA helicase RuvB	K03551	Homologous recombination	cjr03440	AAW35989	335	No	Yes
7	CJE1486	aminotransferase, DegT/DnrJ/EryC1/StrS family	K15895	Amino sugar and nucleotide sugar metabolism	cjr00520	AAW35927	376	No	Yes
8	CJE0667	replicative DNA helicase	K02314	DNA replication	cjr03030	AAW35822	458	No	Yes
9	CJE0684	(di)nucleoside polyphosphate hydrolase	K08311	RNA degradation	cjr03018	AAW35806	156	No	Yes
10	CJE1406	oxidoreductase, 2-nitropropane dioxygenase family	K00459	Nitrogen metabolism	cjr00910	AAW35725	363	No	Yes
11	CJE1401	quinone-reactive Ni/Fe hydrogenase, cytochrome b subunit	K03620	Two-component system	cjr02020	AAW35720	238	No	Yes
12	CJE1382	excinuclease ABC, C subunit	K03703	Nucleotide excision repair	cjr03420	AAW35702	600	No	Yes
13	CJE1343	HD/HDIG/KH domain protein	K18682	RNA degradation	cjr03018	AAW35664	517	No	Yes
14	CJE1292	transcription termination factor Rho	K03628	RNA degradation	cjr03018	AAW35614	432	No	Yes
15	CJE1264	general glycosylation pathway protein	K15910	Amino sugar and nucleotide sugar metabolism	cjr00520	AAW35586	386	No	Yes
16	CJE1228	transcription-repair coupling factor	K03723	Nucleotide excision repair	cjr03420	AAW35550	978	No	Yes
17	CJE1196	DNA mismatch repair protein	K07456	Mismatch repair	cjr03430	AAW35521	735	No	Yes
18	CJE1067	hippurate hydrolase	K01451	Phenylalanine metabolism	cjr00360	AAW35395	383	No	Yes
19	CJE0993	thioesterase family protein	K10806	Biosynthesis of unsaturated fatty acids	cjr01040	AAW35326	137	No	Yes
20	CJE0966	flagellin family protein	K02397	Flagellar assembly	cjr02040	AAW35299	750	No	Yes
21	CJE0951	thiol:disulfide interchange protein DsbA	K03673	Cationic antimicrobial peptide (CAMP) resistance	cjr01503	AAW35288	220	No	Yes
22	CJE0949	para-aminobenzoate synthase glutamine amidotransferase, component I	K03342	Folate biosynthesis	cjr00790	AAW35286	594	No	Yes
23	CJE0943	signal peptidase I	K03100	Protein export	cjr03060	AAW35280	282	No	Yes
24	CJE0892	integral membrane protein MviN	K03980	Peptidoglycan biosynthesis	cjr00550	AAW35229	483	No	Yes
25	CJE0890	Holliday junction DNA helicase RuvA	K03550	Homologous recombination	cjr03440	AAW35227	183	No	Yes
26	CJE0860	flagellar basal-body P-ring formation protein FlgA, putative	K02386	Flagellar assembly	cjr02040	AAW35197	220	No	Yes
27	CJE0514	ATP-dependent DNA helicase RecG	K03655	Homologous recombination	cjr03440	AAW35101	607	No	Yes
28	CJE0410	lipoprotein signal peptidase	K03101	Protein export	cjr03060	AAW34999	156	No	Yes
29	CJE0402	phosphatase, Ppx/GppA family	K01524	Purine metabolism	cjr00230	AAW34991	486	No	Yes

(Continued)

Table 2. (Continued)

No.	KEGG ID	Protein Name	KO Number	Pathway	Pathway ID	NCBI Protein ID	Length (a.a.)	Chokepoint Enzyme	Virulence Factor
30	CJE0390	excinuclease ABC, A subunit	K03701	Nucleotide excision repair	cjr03420	AAW34979	940	No	Yes
31	CJE0288	carbonic anhydrase	K01673	Nitrogen metabolism	cjr00910	AAW34880	211	No	Yes
32	CJE0782	primosomal protein N'	K04066	Homologous recombination	cjr03440	AAW34570	617	No	Yes
33	CJE0771	RNA polymerase sigma-54 factor	K03092	Two-component system	cjr02020	AAW34562	416	No	Yes
34	CJE1897	crossover junction endodeoxyribonucleaseRuvC	K01159	Homologous recombination	cjr03440	AAW34497	160	No	Yes
35	CJE1845	RecA	K03553	Homologous recombination	cjr03440	AAW36267	343	No	Yes
36	CJE0001	chromosomal replication initiator protein DnaA	K02313	Two-component system	cjr02020	AAW34498	440	No	Yes
37	CJE0198	undecaprenol kinase, putative	K06153	Peptidoglycan biosynthesis	cjr00550	AAW34792	267	No	Yes
38	CJE0406	Conserved hypothetical ProteinTIGR00023	K08591	Glycerolipid metabolism	cjr00561	AAW34995.1	202	No	Yes

https://doi.org/10.1371/journal.pone.0198170.t002

Characterization of vaccine and drug targets

Potential target proteins were characterized for various properties such as subcellular localization, antigenicity, functional importance, druggability and involvement in pathways having more than one target protein (Fig 3). Among the 38 targets, 28 are cytoplasmic proteins, 5 inner membrane proteins, 1 periplasmic protein and 1 target has potential to be both cytoplasmic and outer membrane bound. From this result, it is obvious that 2 outer membrane proteins (1 outer membrane and 1 extracellular) are candidate vaccine targets and the other 36 targets may be potential drug targets (Table 3, S13 Table). Utilizing parameters mentioned in the method section, 19 proteins were found as antigenic proteins and the other 19 proteins were considered as non-antigenic. Among the antigenic proteins Flagellin family protein, Lipoprotein signal peptidase, 'Quinone-reactive Ni/Fe hydrogenase, cytochrome b subunit', N-acetylmuramoyl-L-alanine amidase and RecA were considered as the best 5 potential antigens with antigenic scores of 0.6974, 0.6913, 0.6486, 0.6402 and 0.6336 respectively (S14 Table). Among these targets the outer membrane protein N-acetylmuramoyl-L-alanine amidase, KEGG ID: CJE1405, was found with high antigenic score of 0.6402 as well as human non-allergen. Potential epitopes were searched within this protein for peptide based vaccine design approaches. Five (5) potential T-cell epitopes (SSKTLNTNY, IFVFLVFAF, LYTRSSDKF, QKFRYVVSF and DYRLVISQF) were predicted from this antigenic protein as these 9 mer amino acids showed the highest combined score of 12 Major Histocompatibility Comeplex (MHC) supertypes and human non-allergenecity (S15 Table). Therefore, this antigen has high potential to be a viable and effective vaccine target (Table 3, S14 Table). Then we have predicted the MHC-I interacting molecules on the basis of their percentile rank (>10) (S16 and S17 Tables) as the lower percentile rank is considered the strong binding affinity to T-cell epitopes[60]. The epitope conservancy analysis confirmed that these T-cell epitopes are well conserved among the CjR targets (S17 Table). We have also explored the B-cell antigenic properties from 6 different B-cell epitope prediction tools [61–66]. We have identified that the peptide region (302–310) 'NNEKENQKP' might be predicted as B-cell epitope from the all cross-referencing data (S1 Fig). The functional importance of each target of λ (38 nos) list was investigated. λ assuming that the more interactions a protein forms with other proteins, then the importance of that protein in a protein system also increases.
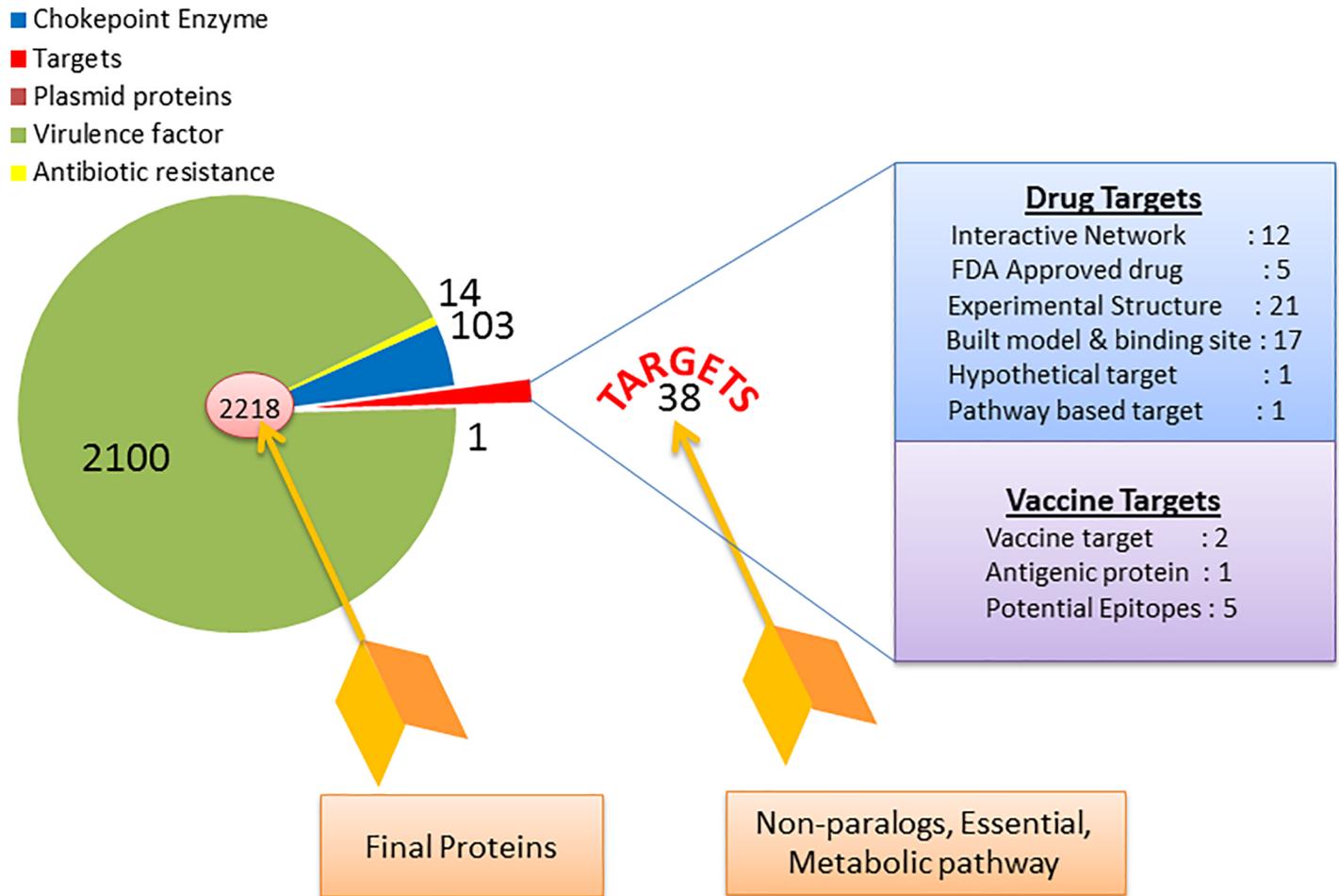
Fig 3. Prioritization of identified therapeutic targets. The nature of targets was categorized on the basis of their properties.

https://doi.org/10.1371/journal.pone.0198170.g003

Common interactions of a protein with other specific proteins among all of the strains of a specific species indicate an evolutionary importance for those interactions. For the species of *C. jejuni*, 12 targets available in STRING 10[31]. CJE1810, CJE1554, CJE0667, CJE1382, CJE1264, CJE0890, CJE0410, CJE0390, CJE0771, CJE1897, CJE1845 and CJE0001 showed interactions with other proteins and these interactions are common to all strains of *C. jejuni*. Among them CJE0890 interacts with four genes ruvB, ruvC, recN, uvrB and it has highest number of interactions. So, CJE0890 can be considered as a superior potential drug target (Table 3, S18 Table). In the current approach, the druggability of the short-listed (38 nos) potential targets was evaluated by a sequence similarity search against targets from DrugBank. Five proteins CJE1596, CJE1554, CJE1486, CJE0993 and CJE0390 were found to show affinity with FDA approved drugs as they have homology with the related proteins (Table 3, S19 Table). Among them CJE0390 is the most eligible druggable target as it interacts with more proteins (Table 3) demonstrating its importance for *CjR*. *CJE0390* is targetable by Colchicine, Silodosin, Etravirine, Doxorubicin, Gramicidin D, Cyclosporine, Adenosine triphosphate, Pravastatin, Fluvoxamine, Fluconazole, Erythromycin, Caffeine and Reserpine as inducer, substrate or inhibitor (S19 Table). The new drugs (synthesize or designed) could emerge by targeting the remaining targets 'novel drug targets' based on drug binding affinity, toxicity measurement and ADME (Absorption, Distribution, Metabolism and Excretion) properties.

Table 3. Characteristic features of pathway based drug targets.

No.	KEGG ID	Subcellular Location Analysis	Antigenicity Analysis	Common Interacting Gene	Druggability
1	CJE1596	Cytoplasmic	0.4297		Druggable
2	CJE1405	Outer Membrane	0.6402		Novel
3	CJE1810	Cytoplasmic	0.2813	2	Novel
4	CJE1654	Cytoplasmic	0.3066		Novel
5	CJE1566	Cytoplasmic	0.4355		Novel
6	CJE1554	Cytoplasmic	0.4604	2	Druggable
7	CJE1486	Cytoplasmic	0.4548		Druggable
8	CJE0667	Cytoplasmic	0.3497	1	Novel
9	CJE0684	Cytoplasmic	0.3053		Novel
10	CJE1406	Cytoplasmic	0.392		Novel
11	CJE1401	Inner Membrane	0.6486		Novel
12	CJE1382	Cytoplasmic	0.3991	2	Novel
13	CJE1343	Cytoplasmic	0.4653		Novel
14	CJE1292	Cytoplasmic	0.3294		Novel
15	CJE1264	Cytoplasmic	0.2672	1	Novel
16	CJE1228	Cytoplasmic	0.3231		Novel
17	CJE1196	Cytoplasmic	0.3961		Novel
18	CJE1067	Cytoplasmic	0.4554		Novel
19	CJE0993	Cytoplasmic, Outer Membrane	0.3742		Druggable
20	CJE0966	Extracellular	0.6974		Novel
21	CJE0951	Periplasmic	0.3157		Novel
22	CJE0949	Cytoplasmic	0.2475		Novel
23	CJE0943	Cytoplasmic Membrane	0.4205		Novel
24	CJE0892	Inner Membrane	0.5468		Novel
25	CJE0890	Cytoplasmic	0.422	4	Novel
26	CJE0860	Cytoplasmic	0.4081		Novel
27	CJE0514	Cytoplasmic	0.3515		Novel
28	CJE0410	Inner Membrane	0.6913	2	Novel
29	CJE0402	Cytoplasmic	0.2989		Novel
30	CJE0390	Cytoplasmic	0.4635	3	Druggable
31	CJE0288	Cytoplasmic	0.2426		Novel
32	CJE0782	Cytoplasmic	0.3231		Novel
33	CJE0771	Cytoplasmic	0.3234	1	Novel
34	CJE1897	Cytoplasmic	0.5022	3	Novel
35	CJE1845	Cytoplasmic	0.6336	2	Novel
36	CJE0001	Cytoplasmic	0.439	3	Novel
37	CJE0198	Inner Membrane	0.3805		Novel
38	CJE0406	Inner Membrane	0.586		Novel

Finally, to design drugs against a pathway, multiple potential pathways were analyzed. Pathways having more target proteins were considered as more potential target pathways. For this, 17 pathways were analyzed. Among them 6 pathways were found to have proteins connected to functions in human pathways and were excluded. Among the remaining 11 pathways, peptidoglycan biosynthesis contains 14 targets, the highest number of target proteins compared to other pathways. Therefore, peptidoglycan biosynthesis was considered as the pathway with most potential for discovering a suitable drug target (S20 and S12 Tables).
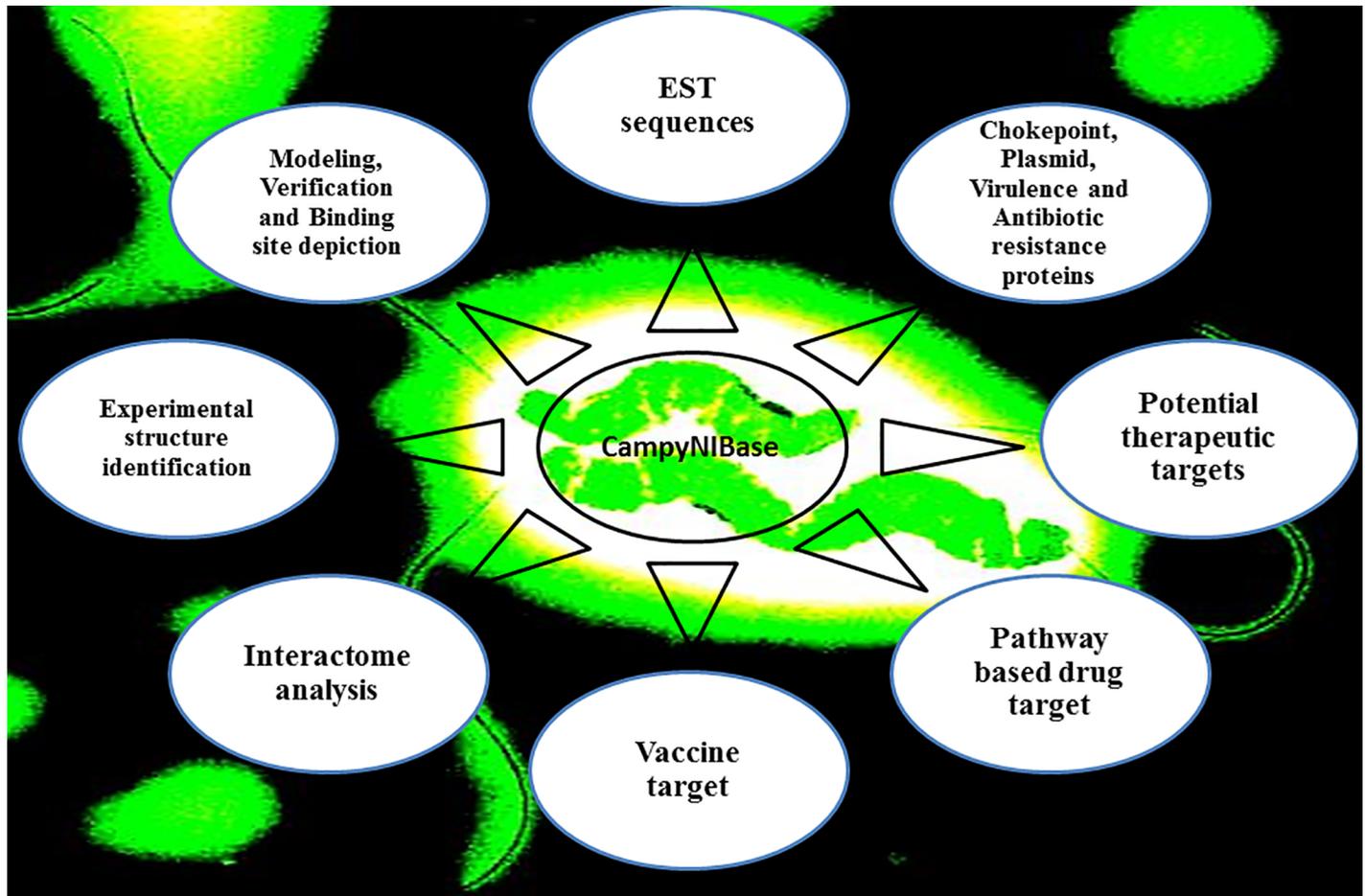
Fig 4. Architecture of 'CampyNIBase' database. The contents of developed database were categorized with different types of important relevant information.

Database development. We have also exploited the experimental and 3D structure of our identified 38 drug targets. We have found 21 experimental structures from the available database. We then built 17 3D structures from the remaining sets of targets. The quality of predicted models was satisfactory, however some of the models showed the favored residues below 90%. Binding sites was also predicted by utilizing Cofactor server.

We have deposited all the potential targets into our databases under the section of Essential proteins and Therapeutic targets. We have also stored the predicted EST sequence of *CjR* strain as the EST sequences are also important for denoting potential epitopes as well as vaccine and drug development (Figs 4 and 5).Furthermore, We have provided the link of our selected targets to be accessible to the other strains including *Campylobacter jejuni* subsp. doylei 269.97, *Campylobacter jejuni* subsp. jejuni 81–176, *Campylobacterjejuni* subsp. jejuni 81116 and *Campylobacter jejuni* subsp. jejuni NCTC 11168 directly via our database (Please refer to utility section of the database).

Discussion

To our knowledge the current subtractive analysis performed in this study is the first report on computational analysis to identify and characterize therapeutic targets of *CjR*. We have selected *CjR* strainof campylobacter which is most prevalent in Bangladesh among the others.
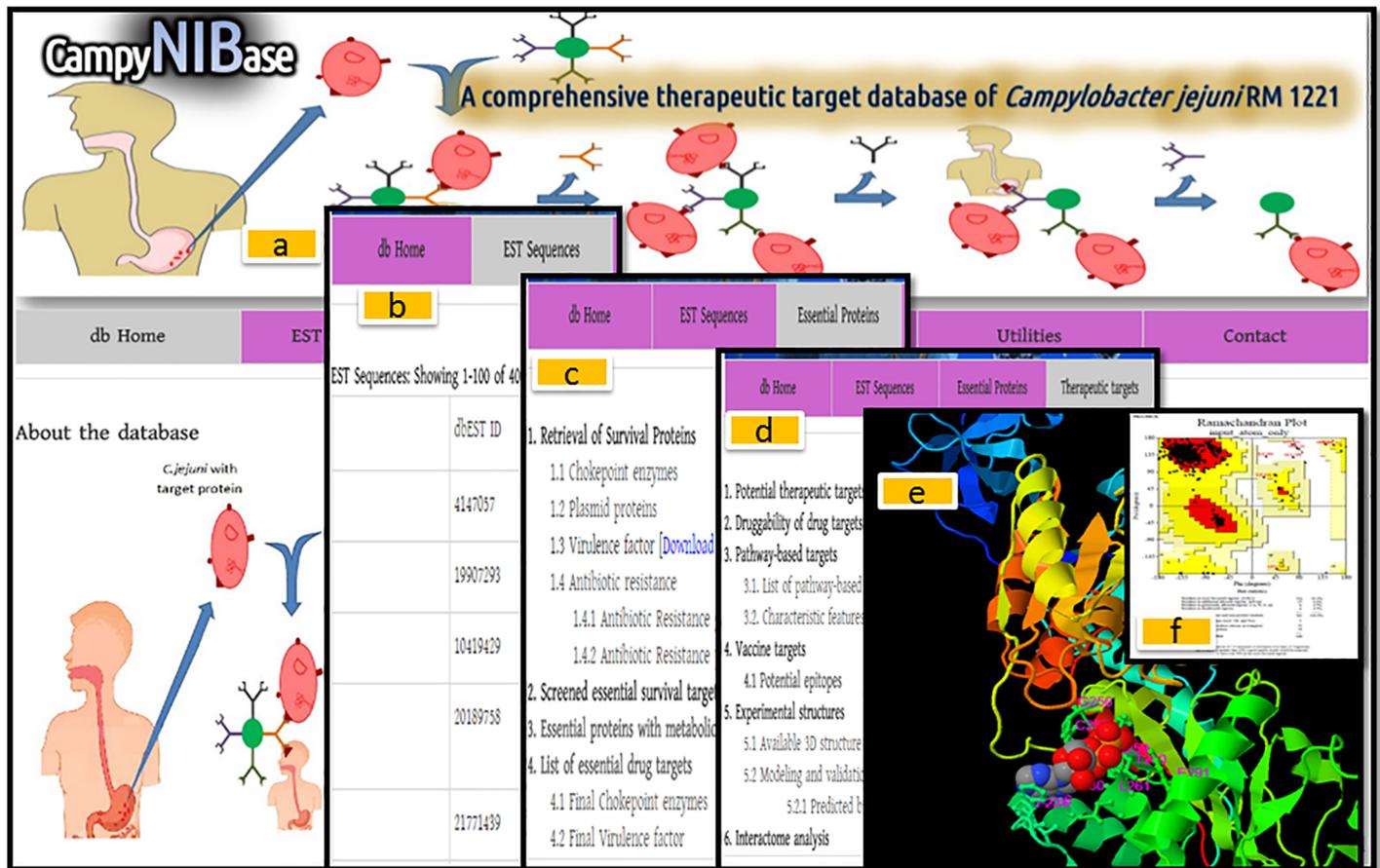
Fig 5. The snapshot of the 'CampyNIBase' database. (a) Homepage tab: summary of *campylobacter jejuni* RM1221. (b) EST tab: EST related information. (c) Essential proteins tab: essential proteins of *campylobacter jejuni* RM1221. (d) Therapeutic targets tab: identified targets and prioritization. (e) Information of built and experimental 3D structure. (f) Model verification.

In molecular level, *CjR* strain is reported as one of the most frequent etiologic agent to cause about 65% diarrheal diseases in different clinical isolates in Bangladesh [74,75]. Besides this, technically to work with *CjR* strain in Bioinformatics analysis is advantageous asthis strain is well annotated with complete genome in KEGG and KAAS database and other bioinformatics resources as well. We have exploited total proteins of *CjR* for pathway based drug design. Whole proteome annotation identified a substantial number (2218) of protein sequences through various databases and literature searches. In this study, we have particularly focused on chokepoint enzymes, plasmid proteins, antibiotic resistance genes which are essential for *CjR*. In a metabolic reaction, only a specific substrate can be consumed and specific product produced during catalysis by a specific enzyme. This type of reaction is known as a chokepoint reaction and the enzyme that catalyzes the reaction is called a chokepoint enzyme. Therefore, certain chokepoint enzymes of a bacterial strain have potential to be a promising drug target [23]. As plasmid proteins can confer unique characteristics to a specific strain, we have collected their sequences through literature survey. In addition, antibiotic resistances by *C. jejuni* are mutations in certain genes. *C. jejuni* strains that have acquired multiple antibiotic resistances (MAR) overexpress *cmeB* compared to normal strains [28]. Many *C. jejuni* strains harbour point mutations in the *gyrA* gene to facilitate their resistance to fluoroquinolone [29]. Moreover, another resistance gene called *aphA-7*, which encodes a kanamycin resistant

phosphotransferase, is found to be native in the genome of numerous *Campylobacter* species [30]. Usually, successful pathogenic bacteria rely on multiple virulent factors for survival and effective replication inside a host which in turn makes them potential drug targets (S1–S5 Tables).

The annotated proteome was filtered for duplicate and smaller protein sequences to avoid any ambiguity and identify ideal proteins respectively. As *CjR* is responsible for food poisoning and diarrhea, the genes responsible may be common to other bacterial species that also cause diarrhea and food poisoning. In addition, these genes can be important for their survival. To sort out these gene products/proteins, we have also filtered the proteins by subjecting them to BLASTp against the other bacterial species responsible for food poisoning, available in DEG 10 database [54]. Many of the successfully identified candidate drug targets revealed that they are involved in various metabolic pathways, essential for pathogen survival and absent in the host organism [76]. We have analyzed the filtered data for these properties and all the homologous human proteins were removed through BLASTp analysis. A further challenge when developing drugs against bacterial proteins is that humans contain a wide variety of bacterial species in their gut microbiome including symbiotic commensal bacteria. The proteins of these beneficial floras should ideally not be targeted. To avoid this problem BLASTp against these floras was also carried out. In this case, we found no protein that matched with the proteins of these floras.

KEGG database [55] contains genes of both complete and incomplete genomes. Genes of complete genomes are annotated by K numbers. For detailed analysis of proteins in KEGG database, annotation is required. To collect annotated proteins, KEGG Automatic Annotation Server (KAAS) [56] was used. Following annotation, we could easily identify proteins involved in pathways common to *CjR* and *H. sapiens*. These common proteins were removed to avoid cross targeting of human pathways by a pathway based drug. Subsequently, proteins belonging to unique pathways of *C. jejuni* were categorized based on the presence of proteins involved in either one or more than one pathway. Proteins found in one pathway were collected to aid in designing of pathway based drugs. Proteins with no involvement in pathway were not considered, as the aim of this study is to estimate pathway based drug targets. Again, proteins involved in more than one pathway were also excluded for the sake of simplicity of designing a pathway based drug. Some proteins were involved in only KEGG module. These are also not eligible as the module is a part of pathway, rather than whole pathway. These all ineligible proteins, other than pathway based drug targets, were named peptide drug targets. We have followed some published work in some reputed journal in which the most antigenic protein was selected for the peptide based epitope design [77–82]. A simple analysis was applied to them. The antigenicity of every peptide drug target was estimated by Vaxijen [59] as they might be helpful for vaccination. We have predicted T-cell and B-cell epitope from the highest probable antigenic target (S1 Fig, S16 and S17 Tables).

Thus, by this systematic subtraction analyses, owing to the considerations for all possible vital parameters, 38 potential therapeutic targets were identified (S6–S12 Tables). We have prioritized some pathway based identified therapeutic targets through characterization for certain properties such as subcellular localization, antigenicity, interactome, drug binding as well as whole pathway targeting ability. These analyses were handful for dividing them into drug or vaccine target. Comparing both subcellular location and antigenic score of all the 38 targets, N-acetylmuramoyl-L-alanine amidase, KEGG ID: CJE1405, was found as having the most potential as a vaccine target (S13–S20 Tables). According to DrugBank 3.0 [67] currently no approved drug or vaccine has been designed against this target and further wet lab validation experiments should be planned. Overall the characterization here revealed 12 drug targets based on interactome analysis, 5 targets to be dealt with FDA approved drugs and 1 for the

pathway based drug target (Fig 3). In interacting genes analysis, it was found that a cytoplasmic protein, CJE0890, interacts with four genes and these interactions are common to all the *C. jejuni* strains available in STRING 10 (Table 3). So, it might become an essential target protein of *CjR* to design a drug against. Another important cytoplasmic protein, CJE0390, shows three interactions and there are available FDA approved drugs that target it (Table 3).

A drug target can be either a single protein or the whole pathway and drugs can be designed to target whole pathway. Pathways containing more eligible target proteins were considered as a target pathway with higher potential. From our analysis, the peptidoglycan biosynthesis pathway was found to contain the highest number of eligible target proteins and considered as the most significant target pathway.

As the *in silico* identified effective therapeutic [83–85], the probable vaccine and drug targets of *CjR* identified through this *in silico* approach are expected to contribute to development of new effective drugs and vaccines to treat campylobacteriosis. Our developed database (Figs 4 and 5) might also enable further exploration and characterization of *CjR* for the development of *CjR* specific therapeutics.

Conclusions

The approach used in this study could corroborate to a powerful channel of analysis with rational accuracy toward identification of important essential genes in bacteria. This study identified one vaccine, 2 drug targets and 1 pathway based drug target. These potential therapeutic targets could be further validated experimentally through the drug and vaccine design pipelines. Together with 3D structures and other data, a comprehensive database 'CampyNIBase' has been developed to assist future *CjR* research.

Supporting information

S1 Table. Final chokepoint proteins.
(XLSX)

S2 Table. Plasmid protein.
(XLSX)

S3 Table. Final virulence factors.
(XLSX)

S4 Table. Final resistant protein.
(XLSX)

S5 Table. Combined total proteins.
(XLSX)

S6 Table. Proteins after duplicate rejection.
(XLSX)

S7 Table. Eligible protein collection.
(XLSX)

S8 Table. Selected filtered by BLASTp against HUMAN.
(XLSX)

S9 Table. Screened essential proteins by DEG.
(XLSX)

S10 Table. Proteins with KO.
(XLSX)

S11 Table. Final drug targets with pathway details.
(XLSX)

S12 Table. Final drug targets.
(XLSX)

S13 Table. Cellular localization.
(XLSX)

S14 Table. Antigenic prediction.
(XLSX)

S15 Table. Potential epitopes.
(XLSX)

S16 Table. T cell and MHC-I molecules.
(XLSX)

S17 Table. T-cell epitopes with epitope conservancy.
(XLSX)

S18 Table. Interactome analysis.
(XLSX)

S19 Table. Druggability.
(XLSX)

S20 Table. Pathway anlysis.
(XLSX)

S1 Fig. Prediction of B-cell antigenic properties. (A) Bepipred linear epitope prediction. (B) Chou and Fasman beta-turn prediction. (C) Emini surface accessibility prediction.(D) Karplus and Schulz flexibility prediction. (E) Kolaskar and Tongaonkar antigenicity prediction. (F) Parker hydrophilicity prediction. The x-axis and y-axis represent the sequence position and corresponding antigenic properties score, respectively. The threshold level was set as default parameter of the server. The regions were shown in yellow color above the threshold value to be predicted as B-cell epitope. All the cross referencing data showed peptide region NNE-KENQKP (302–310) might be predicted as B-cell epitope.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Mohammad Uzzal Hossain, Md. Salimullah.

Data curation: Mohammad Uzzal Hossain, Taimur Md. Omar, Iftekhar Alam, Keshob Chandra Das, A. K. M. Mohiuddin, Chaman Ara Keya, Md. Salimullah.

Formal analysis: Mohammad Uzzal Hossain, Taimur Md. Omar, Chaman Ara Keya.

Investigation: Iftekhar Alam, Keshob Chandra Das, A. K. M. Mohiuddin, Chaman Ara Keya, Md. Salimullah.

Methodology: Mohammad Uzzal Hossain, Taimur Md. Omar.

Project administration: Md. Salimullah.

Resources: Iftekhar Alam, Keshob Chandra Das, A. K. M. Mohiuddin, Chaman Ara Keya, Md. Salimullah.

Software: Taimur Md. Omar, Iftekhar Alam, Keshob Chandra Das, A. K. M. Mohiuddin, Chaman Ara Keya.

Supervision: Md. Salimullah.

Validation: Mohammad Uzzal Hossain, Taimur Md. Omar, Md. Salimullah.

Visualization: Mohammad Uzzal Hossain, Taimur Md. Omar, Md. Salimullah.

Writing – original draft: Mohammad Uzzal Hossain, Taimur Md. Omar, Md. Salimullah.

Writing – review & editing: Mohammad Uzzal Hossain, Md. Salimullah.

References

1. Penner JL (1988) The Genus *Campylobacter*: A decade of progress. *ClinMicrobiol Rev* 1: 157–172

2. Lastovica AJ, Skirrow MB (2000) Clinical significance of Campylobacter and related species other than Campylobacter jejuni and C. coli. In: Nachamkin I, Blaser MJ, editors. Campylobacter. Washington (DC): ASM Press: 89–120.

3. Altekruse SF, Stern NJ, Fields PI, Swerdlow DL (1999) *Campylobacter jejuni* and emerging foodborne pathogen. *Emerg Infect Dis* 5: 28–35. https://doi.org/10.3201/eid0501.990104 PMID: 10081669

4. Nachamkin I BM Allos T Ho (1998) *Campylobacter* species and Guillain-Barre syndrome. *Clin. Microbiol. Rev.* 11:555–567. PMID: 9665983

5. Willison HJ, Hanlon GM (1999) The immunopathogenesis of Miller Fisher *syndrome. J Neuroimmunol.* 100: 3–12 PMID: 10695710

6. Blaser MJ, Berkowitz ID, LaForce FM, Cravens J, Reller LB, Wang WL (1979) *Campylobacter* enteritis: Clinical and epidemiologic features. *Ann Intern Med.* 91: 179–185. PMID: 380433

7. Miller WG, Mandrell RE (2004) Campylobacter in the food and water supply: Prevalence, outbreaks, isolation, and detection. In: Ketley J, Konkel ME, editors. Campylobacter jejuni: New perspectives in molecular and cellular biology. Norfolk (United Kingdom): Horizon Scientific Press. pp. 109–163.

8. Scharff RL (2012) Economic burden from health losses due to foodborne illness in the United States. *J Food Prot.* 75: 123–131. https://doi.org/10.4315/0362-028X.JFP-11-058 PMID: 22221364

9. Heredia N, Wesley I, Garcia S (2009) Microbiologically Safe Foods. In: Hoboken NJ. Foodborne Pathogens and Toxins: Overview. USA: John Wlley& Sons. pp 20–22.

10. Silva J, Leite D, Fernandes M, Mena C, Gibbs PA, Teixeira P. (2011) *Campylobacter* spp. as a foodborne pathogen: A review. *Front Microbiol.* 2: 1–12. https://doi.org/10.3389/fmicb.2011.00001

11. EFSA (2007) The community summary report on trends and sources of zoonoses, zoonotic agents, antimicrobial resistance and foodborne outbreaks in the European Union in 2006. EFSA J 130: 130–155.

12. EFSA (2010) The community summary report on trends and sources of zoonoses, zoonotic agents, and foodborne outbreaks in the European Union in 2008. *EFSA* J8: 1496–1906

13. Altekruse SF, Stern NJ, Fields PI, Swerdlow DL (1999) *Campylobacter jejuni* and emerging foodborne pathogen. *Emerg Infect Dis.* 5: 28–35. https://doi.org/10.3201/eid0501.990104 PMID: 10081669

14. Coker AO, Isokpehi RD, Thomas BN, Amisu KO, Obi CL (2002) Human campylobacteriosis in developing countries. *Emerg Infect Dis.* 8: 237–243. https://doi.org/10.3201/eid0803.010233 PMID: 11927019

15. Allsop AE (1998) Bacterial genome sequencing and drug discovery. *CurrOpinBiotechnol.* 9(6):637–642.

16. Stumm G, Russ A, Nehls M (2002) Deductive genomics: a functional approach to identify innovative drug targets in the post-genome era. *Am J Pharmacogenomics* 2(4): 263–271. PMID: 12421097

17. Damte D, Suh JW, Lee SJ, Yohannes SB, Hossain MA, Park SC (2013) Putative drug and vaccine target protein identification using comparative genomic analysis of KEGG annotated metabolic pathways of Mycoplasma hyopneumoniae. *Genomics*. 102(1):47–56. https://doi.org/10.1016/j.ygeno.2013.04.011 PMID: 23628646

18. Amineni U, Pradhan D, Marisetty H (2010) In silico identification of common putative drug targets in Leptospira interrogans. *J Chem Biol*. 3(4), 165–173. https://doi.org/10.1007/s12154-010-0039-1 PMID: 21572503

19. Barh D, Kumar A (2009) In silico identification of candidate drug and vaccine targets from various pathways in Neisseria gonorrhoeae. *In Silico Biol*. 9(4):225–231. PMID: 20109152

20. Rahman A, Noore S, Hasan A, Ullah R, Rahman H, Hossain A et. al. (2014) Identification of potential drug targets by subtractive genome analysis of Bacillus anthracis A0248: An in silico approach. *ComputBiol Chem*. 52:66–72.

21. Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D et al. (2000) The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature*. 403: 665–668. https://doi.org/10.1038/35001088 PMID: 10688204

22. Gundogdu O, Bentley SD, Holden MT, Parkhill J, Dorrell N, Wren BW (2007) Re-annotation and re-analysis of the *Campylobacter jejuni*NCTC11168 genome sequence. BMC Genomics. 8: 471–472. https://doi.org/10.1186/1471-2164-8-471

23. Yeh I, Hanekamp T, Tsoka S, Karp PD, Altman RB (2004) Computational Analysis of *Plasmodium falciparum*Metabolism: Organizing Genomic Information to Facilitate Drug Discovery. Genome Research. 14(5): 917–24. https://doi.org/10.1101/gr.2050304 PMID: 15078855

24. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA (2004) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*. 42, D459–471.

25. Fouts DE, Mongodin EF, Mandrell RE, Miller WG, Rasko DA, Ravel J (2005) Major Structural Differences and Novel Potential Virulence Mechanisms from the Genomes of Multiple Campylobacter Species. *PLoS Biol*, 3(1): e15. https://doi.org/10.1371/journal.pbio.0030015 PMID: 15660156

26. Chen L, Xiong Z, Sun L, Yang J, Jin Q (2012) VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res*. 40, D641–645. https://doi.org/10.1093/nar/gkr989 PMID: 22067448

27. Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 35: D61–65. https://doi.org/10.1093/nar/gkl842 PMID: 17130148

28. Pumbwe L, Randall LP, Woodward MJ, Piddock LJ (2004) Expression of the efflux pump genes cmeB, cmeF and the porin gene porA in multiple-antibiotic-resistant *Campylobacter jejuni*. J AntimicrobChemother. 54(2): 341–347.

29. Ge B, McDermott PF, White DG, Meng J (2005) Role of Efflux Pumps and Topoisomerase Mutations in Fluoroquinolone Resistance in *Campylobacter jejuni* and *Campylobacter coli*. Antimicrob Agents Chemother 49(8): 3347–3354. https://doi.org/10.1128/AAC.49.8.3347-3354.2005 PMID: 16048946

30. Tenover FC, Gilbert T, O'Hara P (1989) Nucleotide sequence of a novel kanamycin resistance gene, aphA-7, from Campylobacter jejuni and comparison to other kanamycin phosphotransferase genes. *Plasmid*. 22(1): 52–8. PMID: 2550983

31. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. 39: D561–568. https://doi.org/10.1093/nar/gkq973 PMID: 21045058

32. Liu B, Pop M. ARDB–Antibiotic Resistance Genes Database. *Nucleic Acids Res* 37(Database issue), D443–447(2009). https://doi.org/10.1093/nar/gkn656 PMID: 18832362

33. Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics*. 26(5): 680–2. https://doi.org/10.1093/bioinformatics/btq003 PMID: 20053844

34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol*. 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712

35. Granum PE (1990) Clostridium perfringens toxins involved in food poisoning. *Int J Food Microbiol*. 10 (2): 101–11. PMID: 2205267

36. Al-Mazrou YY (2004) Food poisoning in Saudi Arabia. Potential for prevention? *Saudi Med J*. 25(1): 11–4. PMID: 14758371

37. Mead PS, Griffin PM(1998) Escherichia coli O157:H7. *Lancet*. 352 (9135):1207–12. https://doi.org/10.1016/S0140-6736(98)01267-7 PMID: 9777854

38. Schoeni JL, Wong AC (2005) Bacillus cereus food poisoning and its toxins. *J Food Prot*. 68(3): 636–48. PMID: 15771198

39. Food poisoning, listeriosis, and febrile gastroenteritis. (1997)Nutr Rev.55(2): 57–60.

40. Warren BR, Parish ME, Schneider KR (2006) Shigella as a foodborne pathogen and current methods for detection in food. *Crit* Rev Food Sci Nutr. 46(7): 551–67. https://doi.org/10.1080/10408390500295458 PMID: 16954064

41. Le Loir Y, Baron F, Gautier M (2003) Staphylococcus aureus and food poisoning. *Genet Mol Res*. 312 (1): 63–76.

42. Okuyama Y, Yoshida N (2012) Staphylococcal food poisoning and MRSA enterocolitis. *Nihon Rinsho*. 70(8): 1362–5. PMID: 22894073

43. Ollinger-Snyder P, Matthews ME (1996) Food safety: review and implications for dietitians and dietetic technicians. *J Am Diet Assoc*. 96(2), 163–8, 171; quiz 169–70. https://doi.org/10.1016/S0002-8223(96)00046-6 PMID: 8557943

44. Katzenell U, Shemer J,Bar-Dayan Y(2001) Streptococcal contamination of food: an unusual cause of epidemic pharyngitis. *Epidemiol Infect* 127(2): 179–84. PMID: 11693494

45. Todd EC (1997) Epidemiology of foodborne diseases: a worldwide review. *World Health Stat* Q50(1–2): 30–50.

46. Cabassi E, Mori L (1976) Vibrio parahaemolyticus: aetiological agent of food poisoning. *Folia Vet Lat*. 6 (4), 335–54. PMID: 798718

47. Donovan TJ, van Netten P (1995) Culture media for the isolation and enumeration of *pathogenic* Vibrio species in foods and environmental samples. *Int J Food Microbiol* 26(1): 77–91. PMID: 7662520

48. Tauxe RV(1997) Emerging foodborne diseases: an evolving public health challenge. *Emerg Infect Dis*. 3(4), 425–34. https://doi.org/10.3201/eid0304.970403 PMID: 9366593

49. Kalina GP (1982) Third etiological factor of food poisoning—literature review and formulation of the problem. *ZhMikrobiolEpidemiolImmunobiol* 2, 23–9.

50. Smith JL (1999) Foodborne infections during pregnancy. *J Food Prot*. 62(7): 818–29. PMID: 10419281

51. Wadström T, Ljungh A (1991) Aeromonas and Plesiomonas as food- and waterborne pathogens. *Int J Food Microbiol* 12(4), 303–11. PMID: 1854599

52. Velásquez Jones L, Rodríguez RS (1988) Bacterial-induced diarrhoea. *Drugs* 36. Suppl4: 6–17.

53. Rampal P,Hébuterne X,Fosse T (1989) Infectious diarrhea in the adult. *Rev Prat* 39:(29): 2583–9. PMID: 2602888

54. Luo H, Lin Y, Gao F, Zhang CT, Zhang R (2014) DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res*. 42: D574–80. https://doi.org/10.1093/nar/gkt1131 PMID: 24243843

55. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*. 38: D355–360. https://doi.org/10.1093/nar/gkp896 PMID: 19880382

56. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome genome annotation and pathway reconstruction server. *Nucleic Acids Res*. 35, W182–185. https://doi.org/10.1093/nar/gkm321 PMID: 17526522

57. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R (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 PMID: 20472543

58. Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. *Proteins*. 64: 643–651. https://doi.org/10.1002/prot.21018 PMID: 16752418

59. Doytchinova IA, Flower DR(2007)VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 8,4(

60. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM(2005). Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci*. 62:1025–1037. https://doi.org/10.1007/s00018-005-4528-2 PMID: 15868101

61. Kolaskar AS, Tongaonkar PC(1990). A semi-empirical method for prediction of antigenic determinants on protein antigens.*FEBS Lett*. 1276(1–2):172–174.

62. Emini EA, Hughes JV, Perlow DS, Boger J(1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol*. 55(3):836–839. PMID: 2991600

63. Karplus PA, Schulz GE (1985). Prediction of chain flexibility in proteins. *Naturwissenschaften*. 72 (4):212–213.

64. Larsen JE, Lund O, Nielsen M (2006). Improved method for predicting linear B-cell epitopes. *Immunome Res*. 2(1):2.

65. Chou PY, Fasman GD(1978). Empirical predictions of protein conformation. *Annu Rev Biochem*. 47:251–276. https://doi.org/10.1146/annurev.bi.47.070178.001343 PMID: 354496

66. Rini JM, Schulze-Gahmen U, Wilson IA (1992). Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science*. 255(5047):959–965. PMID: 1546293

67. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A(2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res*. 39: D1035–1041. https://doi.org/10.1093/nar/gkq1126 PMID: 21059682

68. Trott O, Olson AJ (2010). AutoDockVina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry*. 31: 455–461. https://doi.org/10.1002/jcc.21334 PMID: 19499576

69. Sussman JL, Lin D, Jiang J, Manning NO, Prilusky J, Ritter O (1998) Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallographica Section D Biological Crystallography*. 54(6): 1078–1084.

70. Wu S, Zhang Y (2007) LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Res*. 35: 3375–338. https://doi.org/10.1093/nar/gkm251 PMID: 17478507

71. Webb B, Sali A (2014) Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., 5(6):1–5.6.32.

72. Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, 26(2): 283–291.

73. Roy A, Yang J, Zhang Y (2012) COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic acids research, gks372.

74. Islam Z, van Belkum A, Cody AJ, Tabor H, Jacobs BC, Endtz HP (2009) Campylobacter jejuni HS:23 and Guillain-Barré syndrome, Bangladesh. Emerg Infect Dis 15: 1315–1317 https://doi.org/10.3201/eid1508.090120 PMID: 19751603

75. Islam Z, Gilbert M, Mohammad QD, Klaij K, Li J,Van Rijs W (2012) Guillain-Barre´ Syndrome-Related Campylobacter jejuni in Bangladesh: Ganglioside Mimicry and Cross-Reactive Antibodies. PLoS ONE 7(8): e43976. https://doi.org/10.1371/journal.pone.0043976 PMID: 22952833

76. Sharma OP, Kumar MS (2016) Essential proteins and possible therapeutic targets of *Wolbachia*endosymbiont and development of FiloBase-a comprehensive drug target database for Lymphatic filariasis. *Sci. Rep*. 6: 19842 https://doi.org/10.1038/srep19842 PMID: 26806463

77. Yasmin T, Akter S, Debnath M, Ebihara A, Nakagawa T, Nabi AHMN (2016). *In silico* proposition to predict cluster of B- and T-cell epitopes for the usefulness of vaccine design from invasive, virulent and membrane associated proteins of *C. jejuni. In Silico Pharmacology*. 4:5. https://doi.org/10.1186/s40203-016-0020-y PMID: 27376537

78. Hossain MU, Omar TM, Oany AR, Kibria KMK, Shibly AZ, Moniruzzaman M et. al (2018) Design of peptide-based epitope vaccine and further binding site scrutiny led to groundswell in drug discovery against Lassa virus. 3 Biotech. 8(2):81. https://doi.org/10.1007/s13205-018-1106-5 PMID: 29430345

79. Hossain MU, Hashem A, Keya CA, Salimullah M (2016) Therapeutics Insight with Inclusive Immunopharmacology Explication of Human Rotavirus A for the Treatment of Diarrhea. Front. Pharmacol. 7:153. https://doi.org/10.3389/fphar.2016.00153 PMID: 27445802

80. Hasan MA, Khan MA, Datta A, Mazumder MHH, Hossain MU (2015) A comprehensive immunoinformatics and target site study revealed the corner-stone towards Chikungunya virus treatment. Mol Immunol. 65(1):189–204. https://doi.org/10.1016/j.molimm.2014.12.013 PMID: 25682054

81. Khan MA, Hossain MU, Rakib-Uz-Zaman SM, Morshed MN (2015) Epitope -based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study. Scandinavian Journal of Immunology. 82: 25–34. https://doi.org/10.1111/sji.12302 PMID: 25857850

82. Oany AR, Ahmad SAI, Hossain MU, Jyoti TP (2015) Highly conserved antigenic epitope regions in RNA dependent RNA polymerase-L of Crimean-Congo haemorrhagic fever virus: insights about novel vaccine. Adv ApplBioinform. Chem. 8: 1–10.

83. Hossain MU, Khan M, Rakib-Uz-Zaman SM, Ali MT, Islam M, Keya CA (2016) Treating diabetes mellitus: Pharmacophore based designing of potential drugs from Gymnemasylvestre against insulin receptor protein. BioMed Research International. Article ID 3187647, 14 pages

84. Hossain MU, Oany AR, Ahmad SA, Hasan MA, Khan MA, Siddikey MA (2016) Identification of potential inhibitor and enzyme-inhibitor complex on trypanothione reductase to control Chagas disease.

Computational Biology and Chemistry. 65: 29–36 https://doi.org/10.1016/j.compbiolchem.2016.10.002 PMID: 27744094

85. Uzzal HM, Zaffar SA, Md OT, Tous ZF, Sara SU, Hossain MJ et al. (2016) Towards finding the linkage between metabolic and age-related disorders using semantic gene data network. Bioinformation. 12 (1): 22–26. https://doi.org/10.6026/97320630012022 PMID: 27212841