



Novel Target Exploration from Hypothetical Proteins of *Klebsiella pneumoniae* MGH 78578 Reveals a Protein Involved in Host-Pathogen Interaction

G. Pranavathiyani, Jyoti Prava, Athira C. Rajeev and Archana Pan*

Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Pondicherry, India

OPEN ACCESS

Edited by:

Jaipalreddy Panga,
Institute for Systems Biology (ISB),
United States

Reviewed by:

Naazneen Khan,
University of California, San Diego,
United States
Saurabh Mishra,
Cornell University, United States

*Correspondence:

Archana Pan
archana@bicpu.edu.in

Specialty section:

This article was submitted to
Molecular Bacterial Pathogenesis,
a section of the journal
Frontiers in Cellular and Infection
Microbiology

Received: 14 January 2020

Accepted: 28 February 2020

Published: 03 April 2020

Citation:

Pranavathiyani G, Prava J, Rajeev AC
and Pan A (2020) Novel Target
Exploration from Hypothetical Proteins
of *Klebsiella pneumoniae* MGH 78578
Reveals a Protein Involved in
Host-Pathogen Interaction.
Front. Cell. Infect. Microbiol. 10:109.
doi: 10.3389/fcimb.2020.00109

The opportunistic pathogen *Klebsiella pneumoniae* is a causative agent of several hospital-acquired infections. It has become resistant to a wide range of currently available antibiotics, leading to high mortality rates among patients; this has further led to a demand for novel therapeutic intervention to treat such infections. Using a series of *in silico* analyses, the present study aims to explore novel drug/vaccine candidates from the hypothetical proteins of *K. pneumoniae*. A total of 540 proteins were found to be hypothetical in this organism. Analysis of these 540 hypothetical proteins revealed 30 pathogen-specific proteins essential for pathogen survival. A motifs/domain family analysis, similarity search against known proteins, gene ontology, and protein-protein interaction analysis of the shortlisted 30 proteins led to functional assignment for 17 proteins. They were mainly cataloged as enzymes, lipoproteins, stress-induced proteins, transporters, and other proteins (viz., two-component proteins, skeletal proteins and toxins). Among the annotated proteins, 16 proteins, located in the cytoplasm, periplasm, and inner membrane, were considered as potential drug targets, and one extracellular protein was considered as a vaccine candidate. A druggability analysis indicated that the identified 17 drug/vaccine candidates were “novel”. Furthermore, a host-pathogen interaction analysis of these identified target candidates revealed a betaine/carnitine/choline transporters (BCCT) family protein showing interactions with five host proteins. Structure prediction and validation were carried out for this protein, which could aid in structure-based inhibitor design.

Keywords: *Klebsiella pneumoniae*, hypothetical protein, functional annotation, drug target, vaccine target, host-pathogen interaction

INTRODUCTION

Klebsiella pneumoniae is a Gram-negative, encapsulated, non-motile bacterium belonging to the *Enterobacteriaceae* family. It is commonly present in soil, water, and animals, including humans. This organism is a part of normal gut flora in human where it does not cause any infection. However, in healthcare environments, it can colonize in medical devices (viz., ventilators and intravenous catheters) and opportunistically infect immunocompromised patients admitted in

the intensive care unit (CDC.gov., 2020). Indeed, this bacterium causes several infections, such as urinary tract infection, bacteraemia, pneumonia, and liver abscesses in hospitalized patients (Chung, 2016). Patients infected with *Klebsiella* can transmit the pathogen via direct contact or indirectly through contaminated medical devices (CDC.gov., 2020). This opportunistic pathogen is able to form biofilms in various biotic and abiotic surfaces like other pathogenic bacteria, including *Pseudomonas aeruginosa*, *Escherichia coli*, and *Acinetobacter baumannii* (Vuotto et al., 2014; de Campos et al., 2016; Riquelme et al., 2018). The formation of biofilms assists the pathogen in withstanding the host defense mechanism and antimicrobial agents.

Several outbreaks of *K. pneumoniae* infections have been reported in hospital settings from different countries, including China, Israel, Poland, Italy, Colombia, and the United States (Kaye and Dhar, 2016; Ocampo et al., 2016; Baraniak et al., 2017; Krapp et al., 2018; Sotgiu et al., 2018; Liu and Guo, 2019). Reportedly, *K. pneumoniae* accounts for ~12% of all hospital-acquired pneumonia in the world (Ashurst and Dawson, 2019). In addition to the presence of the beta-lactamase enzyme, which makes *K. pneumoniae* antibiotic resistant, an alteration in the upregulation of efflux pumps is reportedly making this opportunistic pathogen resistant to multiple drugs, including the last-resort treatment regimen carbapenems. This leads to high mortality rates among the patients (~50%) (Xu et al., 2017). According to the Center for Disease Control and Prevention (CDC), ~80% of the reported carbapenem-resistant Enterobacteriaceae infections in 2013 were due to *K. pneumoniae* (Ashurst and Dawson, 2019). Thus, the present scenario demands the development of novel therapeutic intervention for treating such bacterial infections.

In many organisms, the molecular functions of more than 30% proteins are unknown; these proteins are termed as “hypothetical proteins”. The functional annotation of hypothetical proteins can enable us to understand their roles in different metabolisms as well as to identify previously unexplored drug targets in an organism (Shahbaaz et al., 2016). Several bioinformatics resources, such as databases and tools, are available for functional annotation of hypothetical proteins. These resources have been successfully used to annotate the functions of hypothetical proteins in different bacterial pathogens, including *Borrelia burgdorferi* (Hassan et al., 2016), *Chlamydia trachomatis* (Turab Naqvi et al., 2017), *Helicobacter pylori* (Naqvi et al., 2016), *Haemophilus influenzae* (Shahbaaz et al., 2013), *Mycobacterium*

tuberculosis (Yang et al., 2019), *Vibrio cholerae* (Islam et al., 2015), and *Staphylococcus aureus* N315 (Prava et al., 2018). Out of the available proteome of *K. pneumoniae* MGH 78578, ~11% is made up of HPs, which can be potential resources to be studied both functionally and structurally. Although bioinformatics studies on a few hypothetical proteins, such as KPN_00953(YcbK) (Teh et al., 2014), KPN_02809 (a Zinc-Dependent Metalloprotease) (Wong et al., 2012), and KPN_00728, KPN_00729 (Chain C and D of Succinate Dehydrogenase, respectively) (Choi et al., 2009), are available, mining and analysis of all hypothetical proteins to shortlist drug/vaccine targets in this pathogen is yet unexplored. In the present study, a series of *in silico* analyses of 540 hypothetical proteins encoded by the *K. pneumoniae* genome were carried out to explore novel drug/vaccine candidates in this organism. The annotated target proteins can be further utilized to design and develop novel inhibitors for the treatment of *Klebsiella* infections.

MATERIALS AND METHODS

Sequence Retrieval

The whole proteome of *K. pneumoniae* subsp. *pneumoniae* MGH 78578 (NC_009648.1) was retrieved from the National Center for Biotechnology Information (NCBI), a comprehensive web portal providing access to genomic and biomedical information (Sayers et al., 2019). The hypothetical proteins (HPs) from the whole proteome of *K. pneumoniae* were obtained using an in-house Perl script (<https://gist.github.com/pranavathiyani>).

Essentiality and Non-homology Analysis

The collected HPs of *K. pneumoniae* were subjected to a similarity search using protein BLAST (BLASTp) (Altschul, 1990) against the essential proteins of bacteria present in the database of essential genes (DEG 15.2) with an *e*-value ≤ 0.0001 and bit-score ≥ 100 as cut-off (Jadhav et al., 2014). DEG is a repository of experimentally determined essential genes of various bacteria, archaea, and eukaryotes (Luo et al., 2014). The query genes/proteins that showed a similarity with genes/proteins of DEG were regarded as possible essential genes or proteins. The HPs that had at least one hit in DEG-BLAST were considered as essential proteins in the current study. Furthermore, a similarity search using BLASTp was carried out between identified essential proteins and the human proteome with a cut-off *e*-value ≥ 0.0001 (Jadhav et al., 2014). The essential pathogen proteins that showed no hit with the human proteome were considered as non-homologous to human proteins and used for further analysis.

Function Prediction

Function prediction of essential non-homologous (ENH) proteins of *K. pneumoniae*, which resulted from the previous analysis, was carried out using bioinformatics resources like InterPro (Mitchell et al., 2019), Pfam (Finn et al., 2016), and NCBI-BLASTp (Altschul, 1990). The ENH protein sequences were submitted to InterPro and Pfam with default parameters to identify their motifs/domain families. InterPro is an integrated online resource of protein databases that provides detailed

Abbreviations: BCCT, Betaine/Carnitine/Choline Transporters; BLAST, Basic Local Alignment Search Tool; CDC, Centers for Disease Control and Prevention; CELLO, subCELLular LOcalization predictor; CELLO2GO, subCELLular LOcalization prediction with functional Gene Ontology; DEG, Database of Essential Genes; ECP, *E. coli* common pilus; ENH, Essential Non-Homologous; GO, Gene Ontology; GRAVY, Grand Average of Hydropathicity; HMM, Hidden Markov Model; HPIDB, Host-Pathogen Interaction Database; HPs, Hypothetical Proteins; NCBI, National Center for Biotechnology Information; PATRIC, Pathosystems Resource Integration Center; PDB, Protein Data Bank; PPI, Protein-protein Interactions; PTS, Phosphotransferase System; SAVES, Structure Analysis and Verification Server; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; SVM, Support Vector Machine; VFDB, Virulence Factor Database.

information about the protein families, domain, and motifs. Pfam uses a Hidden Markov Model-based method to identify the domain families of the proteins. Protein BLAST (BLASTp) search was performed for the ENH proteins to identify homologous sequences with known functions from the NCBI protein database (non-redundant). Moreover, the predicted function of the ENH proteins from InterPro, Pfam, and NCBI-BLASTp search was cross-checked with the function of DEG-hits obtained from an essentiality analysis. The predicted function was also compared and verified using gene ontology (GO) analysis and a protein-protein interaction analysis. The GO analysis was performed by submitting the annotated ENH proteins to CELLO2GO and GO FEAT. CELLO2GO is a web server that performs a similarity search (BLAST) for a given protein sequence to obtain its homologous sequences with GO annotation (molecular function, biological process, and cellular location) (Yu et al., 2014). GO FEAT is an online platform for functional annotation for genomic as well as transcriptomic data based on similarity search (Araujo et al., 2018). Furthermore, the ENH proteins were given as query to STRING database (version 11.0) with medium confidence (0.40) in order to identify functions based on the homolog hits and the interactions among the proteins. STRING database is an integrated resource of experimental and predicted protein-protein interactions (PPI). Currently, STRING comprises more than 2,000 million interactions of 24.6 million proteins from 5,090 organisms (Szklarczyk et al., 2019).

Physicochemical Characterization and Virulence Prediction

The physicochemical properties, such as molecular weight, theoretical isoelectric point (pI), instability index, aliphatic index, and grand average of hydropathicity (GRAVY), of the annotated ENH proteins were computed using the ProtParam tool of ExPasy. ExPasy is a Swiss Institute of Bioinformatics resource portal for tools and databases used in diverse areas of life sciences, including genomics, proteomics, genetics, systems biology, molecular evolution, and transcriptomics (Gasteiger et al., 2005). Virulence proteins among the annotated ENH proteins were identified using a similarity search against the core dataset of virulent proteins from Virulence Factor Database (VFDB) with a cut-off e -value ≤ 0.0001 . VFDB is a database for bacterial virulence factors primarily curated from the scientific literature (Chen et al., 2016). The annotated ENH proteins were submitted to VICMpred, an SVM-based functional classification server for Gram-negative bacterial proteins that functionally classify the proteins into different categories based on amino acid composition (Saha and Raghava, 2006). Further, an MP3 tool was used for the identification of pathogenic proteins among the functionally annotated ENH proteins. The MP3 tool utilizes an integrated approach based on SVM-HMM in order to better provide efficiency and accuracy in predicting pathogenic proteins (Gupta et al., 2014).

Druggability and Subcellular Localization Analysis

To assess the druggability of the annotated ENH proteins, DrugBank and ChEMBL databases were utilized. DrugBank is a comprehensive bioinformatics database for cheminformatics

comprising detailed information about drugs and the corresponding targets (Wishart et al., 2018). ChEMBL is a curated database of bioactive chemical compounds maintained by the EMBL. It includes manually curated data from the scientific literature on drug-like compounds along with their bioactivity determined based on assays (Gaulton et al., 2012). A similarity search was performed between the annotated ENH proteins and known targets of DrugBank and ChEMBL with a cut-off e -value ≤ 0.00001 . Subcellular localization of the proteins was predicted using CELLO (v.2.5), a multi-class SVM-based classification server which uses amino acid sequence features for predicting subcellular localization. In the case of Gram-negative bacteria, the average accuracy of CELLO in localization prediction is 89% (Chen et al., 2006).

Prediction of Host-Pathogen Interactions

The host-pathogen interactions of the annotated ENH proteins were predicted using the interlog method. This method relies on a homology search of the query sequence against the known host-pathogen interaction data. If the pathogen protein “A” interacts with host protein, “B” and if a protein “X” is homologous to protein “A”, then there is a high chance that protein “X” will also interact with protein “B” (Yu et al., 2004). Based on this principle, a homology search of the annotated ENH proteins was performed against the full database of HPIDB with default parameters (identify >50%, query coverage >50% and e -value = 0.00001) to obtain the proteins that interact with the human host. HPIDB 3.0 is a comprehensive database of curated host-pathogen interaction data (Kumar and Nanduri, 2010).

Structure Prediction

The annotated ENH proteins were searched against the Protein Data Bank (PDB) using PSI-BLAST (Altschul, 1990) from NCBI. PDB is a repository of three-dimensional (3D) structure data for different biological macromolecules (Berman et al., 2000). Determining the 3D structure of the protein is important for understanding its molecular function at the atomic level and it also facilitates the process of structure-based drug design. Herein, the structure prediction for the selected protein was carried out using an I-TASSER server (Yang and Zhang, 2015) owing to the low structural similarity of the available 3D structure. The predicted structure was validated using SAVES server (<https://servicesn.mbi.ucla.edu/SAVES/>). The final 3D structure was visualized using the Pymol tool (version 2.3.3 Schrödinger, LLC).

RESULTS AND DISCUSSION

In the present study, HPs encoded by the *K. pneumoniae* MGH 78578 genome were analyzed using the *in silico* approach to shortlist proteins that can be potential drug and vaccine targets. The workflow adopted in the current study is illustrated in **Figure 1**.

Identification of Essential Pathogen-Specific Proteins

The complete genome of *K. pneumoniae* subsp. *pneumoniae* MGH 78578 is a single circular chromosome of 5.31 Mb in size. The genome comprises 5,115 protein-coding genes. The

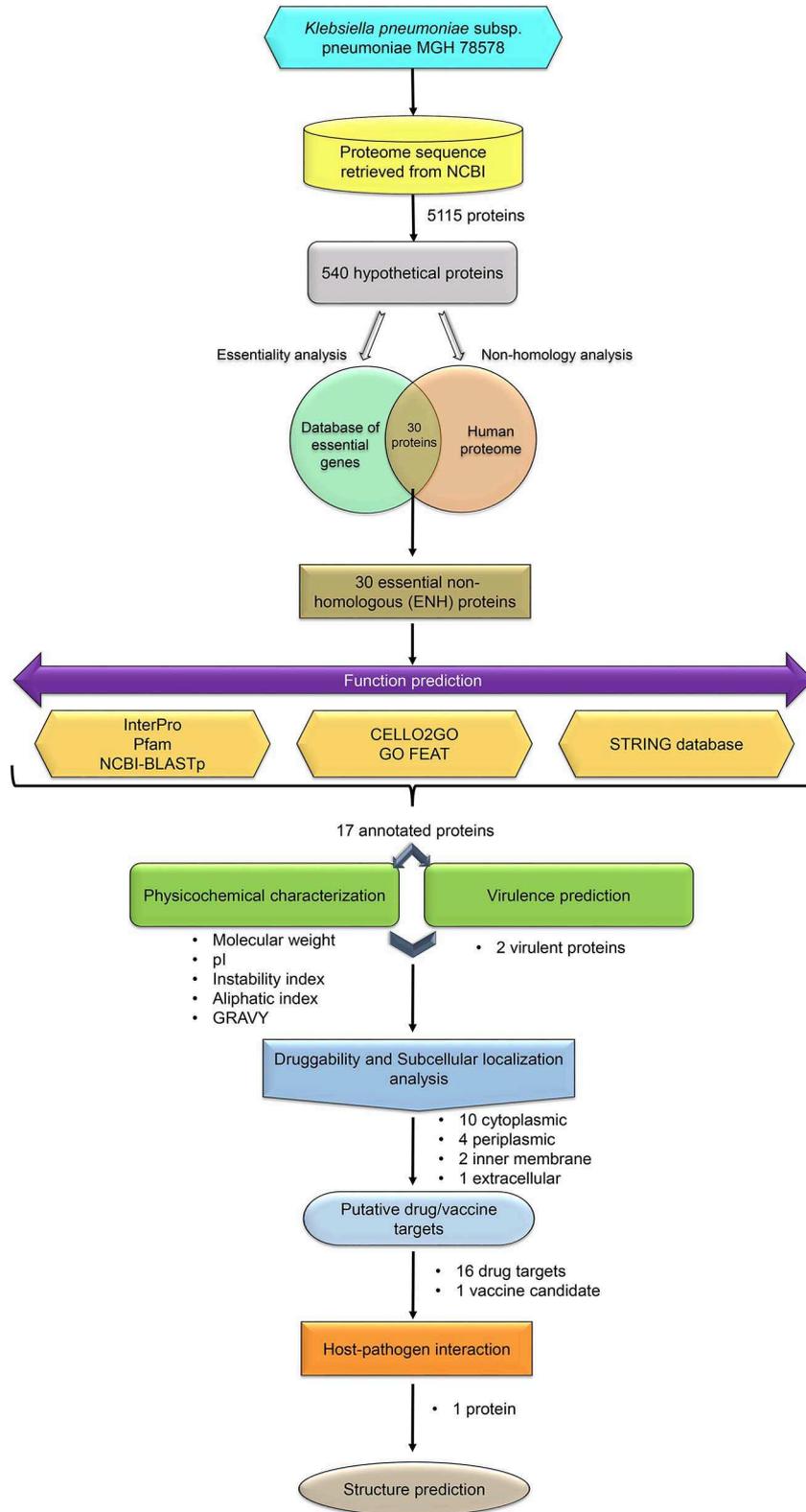


FIGURE 1 | The complete workflow adopted in the present study.

TABLE 1 | List of annotated essential non-homologous proteins and their predicted functions using InterPro, Pfam, and NCBI-BLASTp.

Sl. No	Protein ID	InterPro	Pfam		NCBI BLASTp
			Identifier	Description	
1	WP_002888808.1	Chaperone lipoprotein, PulS/OutS-like (IPR019114)	T2SS_PulS_OutS	Type II secretion system pilotin lipoprotein (PulS_OutS)	Chaperone lipoprotein YacC
2	WP_004222859.1	Endopeptidase, NLPC/P60 domain (IPR000064)	NLPC_P60	NlpC/P60 family	Lipoprotein NlpC/P60 family protein
3	WP_002890061.1	PapD-like superfamily (IPR008962)	PapD_N	Pili and flagellar-assembly chaperone, PapD N-terminal domain	Pilus assembly protein
4	WP_002890284.1	Pyrimidine/purine nucleoside phosphorylase (IPR009664)	DUF1255	Protein of unknown function (DUF1255)	Pyrimidine/purine nucleoside phosphorylase
5	WP_004151327.1	Pheromone shutdown, TraB (IPR002816)	TraB	TraB family	Conjugal transfer protein TraB
6	WP_012068456.1	YbfN-like lipoprotein (IPR025727)	YbfN	YbfN-like lipoprotein	Lipoprotein ybfN
7	WP_004176857.1	Cyclophilin-like domain superfamily (IPR029000)	CT_C_D	Carboxyltransferase domain, subdomain C and D	5-oxoprolinase subunit PxpB
8	WP_023288894.1	–	YmgB	Biofilm development protein YmgB/AriR	Two-component-system connector protein YcgZ
9	WP_073549749.1	BCCT transporter family (IPR000060)	BCCT	BCCT, betaine/carnitine/choline family transporter	BCCT family transporter
10	WP_002898708.1	Stress-induced protein, KGG, repeat (IPR019626)	KGG	Stress-induced bacterial acidophilic repeat motif	Stress-induced protein
11	WP_004150795.1	Stress-induced protein, KGG, repeat (IPR019626)	KGG	Stress-induced bacterial acidophilic repeat motif	Stress-induced acidophilic repeat motif-containing protein
12	WP_041937616.1	Putative bacterial toxin ydaT superfamily (IPR037042)	YdaT_toxin	Putative bacterial toxin ydaT	Bacteriophage regulatory protein CII
13	WP_004143718.1	Stress-induced protein, KGG, repeat (IPR019626)	KGG	Stress-induced bacterial acidophilic repeat motif	Stress-induced acidophilic repeat motif-containing protein
14	WP_002911528.1	Sugar-phosphate isomerase, RpiB/LacA/LacB family (IPR003500)	LacAB_rpiB	Ribose/Galactose Isomerase	Galactose-6-phosphate isomerase LacB subunit
15	WP_002914983.1	Phosphotransferase system, lactose/cellobiose-type IIA subunit (IPR003188)	PTS_IIA	PTS system, Lactose/Cellobiose specific IIA subunit	PTS system cellobiose-specific transporter subunit IIA
16	WP_002918629.1	Barstar-like superfamily (IPR035905)	Barstar	Barstar (barnase inhibitor)	barnase inhibitor
17	WP_015959101.1	Bactofilin A/B (IPR007607)	Bactofilin	Polymer-forming cytoskeletal	Ccm domain-containing protein

corresponding protein sequences were retrieved for the analysis. Out of 5,115 proteins, 540 were found to be HPs in this organism (**Supplementary File 1**) (at the time of retrieval). The identified HPs were subjected to essentiality and non-homology analysis in order to find out essential pathogen-specific proteins. The idea behind the essentiality analysis is to shortlist proteins that are crucial for the survival of the pathogen, and targeting those proteins would thus be lethal for the pathogen. The non-homology analysis will result in proteins that are exclusively present in the pathogen and absent in the human host. This approach can aid in designing drugs that target only pathogen proteins without interfering host system, thereby minimizing side-effects.

The present study utilized experimentally determined essential proteins of 48 bacterial species deposited in DEG to predict essential proteins among *K. pneumoniae* HPs. It is believed that the HPs of bacteria that are similar to known

essential proteins of DEG are possible essential proteins. Out of 540 HPs, 40 were found to be similar to known essential proteins and were thus considered as possible essential proteins of this pathogen. Subsequently, the non-homology analysis of the 40 essential proteins revealed that 30 proteins did not have any hit with the human proteome, and they were therefore non-homologous to humans. These 30 essential non-homologous (ENH) proteins (**Supplementary File 2**) are essential for the survival of the pathogen and are pathogen-specific (present in the pathogen but absent in the human). Theoretically, all the shortlisted ENH proteins have the potential to become good drug/vaccine target(s) in *K. pneumoniae*. However, insight into the functional annotation, physicochemical properties, virulence, druggability, and subcellular localization of these 30 shortlisted proteins would provide an additional layer of refinement for target identification, and those analyses were thus performed subsequently.

Function Prediction of ENH Proteins

The function prediction of 30 ENH proteins using bioinformatics secondary databases and tools, such as InterPro, Pfam, and NCBI BLASTp yielded 17 annotated proteins (**Table 1**) along with 10 uncharacterized domain proteins with unknown functions and three proteins with no hits/similarity (**Supplementary File 3**). It is worth mentioning at this point that our previous work on functional annotation of essential HPs from *Staphylococcus aureus* N315 assessed the performance of these tools using receiver operating characteristic (ROC) curve analysis and was noticed to have annotated the function of HPs with high accuracy (Prava et al., 2018). In the present study, the majority of the 17 annotated proteins fell into the category of enzymes, lipoproteins, transporters, and stress-induced proteins among others. The remaining 10 ENH proteins either belonged to uncharacterized domain(s)/family proteins (YaeP family protein, YccJ like protein, YfdX like protein) or different families of hypothetical proteins (YheO like PAS domain protein). The previous studies of these uncharacterized proteins have suggested their essentiality in bacteria, and some are reported to have condition-based expression in several organisms (Goodacre et al., 2014). For instance, the YfdX protein, which has been reported to be present in several bacterial pathogens, including *Escherichia coli*, *Salmonella enterica* serovar Typhi, and *Typhimurium*, is under the control of a regulator protein, EvgA, responsible for environmental stress expressions (McClelland et al., 2001; Masuda and Church, 2002). A recent study on *Salmonella* infection has demonstrated that YfdX is involved in virulence, antibiotic susceptibility, and in modulating pathogens' growth/survival strategies (Lee et al., 2019). The hypothetical protein WP_002920130.1 with a YheO-like PAS domain was predicted to have putative DNA binding activity apart from its transcription regulatory activity (InterPro entry: IPR013559).

Predicted functions of 17 proteins were verified by a DEG-BLAST search that yielded similar function annotations in other organisms, thereby supporting our results. Of the 17 annotated proteins, eight proteins were found to have orthologs in *Escherichia coli*, a closely related organism of *Klebsiella*; orthologs of the remaining annotated proteins were found to be present in other bacterial species, including *Mycoplasma pulmonis*, *Salmonella*, *Pseudomonas aeruginosa*, and *Haemophilus influenzae* (**Supplementary File 4**). A GO analysis of the annotated 17 proteins revealed that these proteins were involved in various molecular functions, namely, peptidase activity, hydrolase activity, isomerase activity, enzyme regulator activity, ion binding, transmembrane transporter activity, and kinase activity (**Supplementary File 5**). This is in accordance with the function prediction results obtained from InterPro, Pfam, and NCBI BLASTp searches, signifying the reliability of our annotation. Catabolic processes, protein folding, cell differentiation, transport, response to stress, and small molecule metabolic processes were the major biological processes in which these annotated proteins were found to be involved. In addition, the PPI analysis of the 17 annotated proteins mapped 14 proteins in the interaction data, and, of these, 11 proteins were found to have annotated functions that are consistent with our function prediction results (**Supplementary File 6**). In the PPI network, it

was observed that the PTS system protein (cellobiose-specific IIA component) had interacted with 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase, and the secretion protein (chaperone lipoprotein YacC) had interacted with the TraB protein. The detailed discussion on the predicted molecular functions of the 17 annotated ENH proteins can be found in the subsequent section under different functional categories.

Enzymes

Enzymes are a class of proteins that are involved in biochemical reactions in the form of catalysts to convert the substrate(s) to product(s). Among 17 annotated ENH proteins, four belong to different enzyme classes. For instance, the protein WP_002890284.1 was found to be involved in pyrimidine/purine nucleoside phosphorylase activity. In the nucleoside salvage pathway, phosphorylase enzymes participate in catalyzing the reversible phosphorolytic cleavage of the glycosidic bond of pyrimidine/purine nucleosides. They also play an important role in activating prodrugs (as analogs) and as inhibitors for antiparasitic and anticancer agents (Bzowska, 2015). The protein WP_004176857.1 was found to harbor a cyclophilin-like domain, which is involved in cyclosporine A (an immunosuppressive drug) binding. The proteins belonging to this family comprise of a beta-barrel domain, which is the core of cyclophilin-type peptidyl-prolyl cis-trans isomerases activity. This domain speeds up the protein folding by catalyzing the cis-trans isomerization of prolineimide peptide bonds in oligopeptides (Takahashi et al., 1989). The protein WP_002911528.1 is a sugar phosphate isomerase, which includes sugar isomerases like ribose 5-phosphate isomerase B (RpiB), galactose isomerase subunit A (LacA), and galactose isomerase subunit B (LacB). The enzyme galactose-6-phosphate isomerase is induced by the presence of galactose or lactose in the cell. RpiB has a Rossmann-type alpha/beta/alpha sandwich topology and forms a homodimer (Takahashi et al., 1989; Zhang et al., 2003b). This protein is involved in catalyzing the interconversion of D-ribose 5-phosphate and D-ribulose 5-phosphate in the non-oxidative branch of the pentose phosphate pathway (Zhang et al., 2003a). The protein WP_002914983.1 was identified to be involved in the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS). The role of PTS is to serve as a transport system for carbohydrate in bacteria. It takes part in catalyzing the phosphorylation of incoming sugar substrates along with its translocation across the cell membrane, making PTS a link between the uptake and metabolism of sugars (Postma et al., 1993).

Lipoproteins

Bacterial lipoproteins are a group of membrane proteins that are involved in diverse functions, such as cellular physiology, cell division, virulence, adhesion to host cells, and virulence factor translocation. The proteins WP_002888808.1, WP_004222859.1, and WP_012068456.1 were found to be similar to the chaperone lipoprotein YacC, lipoprotein NlpC/P60 family protein, and lipoprotein YbfN family, respectively. Apart from the novel addition of protein YacC, the Chaperon lipoprotein family, PulS/OutS-like comprises of pullulanase secretion protein (PulS)

in *K. pneumoniae* (UniProt ID: P20440), the lipoprotein OutS protein of *Erwinia chrysanthemi* (UniProt ID: Q01567), and a functionally uncharacterized type II secretion protein EtpO (UniProt ID: Q7BSV3) of *E. coli* O157: H7. Reportedly, PulS and OutS interaction facilitates the insertion of secretins into the outer membrane indicating its chaperone-like role in bacterial systems (InterPro entry: IPR019114). In various bacterial lineages, NlpC/P60 proteins categorically belong to the cell wall peptidases family, which hydrolyses the d- γ -glutamyl-meso-diaminopimelate or N-acetylmuramate-L-alanine linkage in the cell wall (Xu et al., 2009).

Stress-Induced Proteins

The present study predicted three proteins (WP_002898708.1, WP_004150795.1, and WP_004143718.1) as stress-induced proteins with a highly conserved KGG repeat. In *E. coli*, YciG protein from yciGFE operon was reported to have a similar motif. The protein YciG, under the regulation of the general stress response controller RpoS, showed significant resistance to thermal and acid stress (Robbe-Saule et al., 2007).

Transporters

Among 17 annotated proteins, two (WP_004151327.1 and WP_073549749.1) were predicted as transporters. The former protein was predicted as a conjugal transfer protein TraB. TraB consists of nucleotide-binding motifs, suggesting its potential energy-providing role in plasmid DNA/Tra protein transport (Chandler and Dunny, 2004). The later one was predicted as an inner-membrane metabolism molecule, and it showed similarity with BCCT family transporter protein (InterPro entry: IPR000060). BCCT represents Betaine/Carnitine/Choline Transporters that have 12 transmembrane regions and four conserved tryptophan in the central region. This protein family is specific to compounds that have quaternary nitrogen atoms (Ziegler et al., 2010).

Toxin, Skeletal, and Two-Component System Proteins

The protein WP_041937616.1 belongs to a putative bacterial toxin YdaT superfamily, which corresponds to a toxin-antitoxin protein system. These genetic modules were found in plasmids as well as in the chromosomes, and they encode for a toxin and its cognate antidote. They are reported to be important in maintaining multi-resistant plasmids and in the evolution of antibiotic resistance (Yamaguchi and Inouye, 2011). There are several theories proposed to explain this plasmid stabilizing toxin-antitoxin systems based on evolution. Targeting the toxin or the anti-toxin would lead to accumulation of proteins, which can lead to lethality.

The bactofilin A/B (WP_015959101.1) family of proteins covers a diverse range of functional roles in cytoskeletal polymer formation, which is conserved among bacterial species. The unique subcellular distribution, and the dynamics of bactofilins in different bacterial species, suggests its roles as versatile structural elements adopting a range of cellular functions (Kühn et al., 2010).

WP_023288894.1 was predicted as YcgZ, a two-component-system connector protein that plays a major role in biofilm formation and in providing an additional input signal into the two-component signaling pathway. YcgZ is a substrate of Lon protease that regulates the expression of an outer membrane protein, OmpF which serves as a passive diffusion pore (Duval et al., 2017). Soo et al. (2011) has reported that YcgZ is associated with resistance against several antibiotics.

Other Proteins

The protein WP_002918629.1 was predicted as a barstar-like superfamily protein. Barstar proteins are small single-chain proteins that tackle the lethal effect of the active barnase enzyme, which is an extracellular ribonuclease. Barstar inhibits the activity of this enzyme by sterically blocking the active site with a helix and adjacent loop segment (Hartley, 1988; Buckle et al., 1994). Another protein WP_002890061.1, which was identified to have PapD-like superfamily domain, acts as a chaperone protein during pili and flagellar assembly. This protein, reportedly found in several pathogenic bacteria, also helps in mediating host cell surface adhesion (Barnhart et al., 2000).

Physicochemical Characterization and Virulence Prediction

In the present study, the molecular weight of the annotated 17 proteins was found to be ranging from 5,953.2 to 35,235.6 Da. The highest and the lowest theoretical isoelectric points (pIs) of the proteins were 10.15 and 4.28, respectively. Molecular weight and pI help in experimental setup for protein purification and crystallization procedures. The instability index measures the stability of the protein in a test tube. If the instability index of a protein is <40 then it is believed to be a stable protein (Gill and von Hippel, 1989). Among the annotated ENH proteins, the index of 11 proteins was found to be <40, and they are thus likely to be stable proteins. The aliphatic index of proteins is calculated on the basis of the number of aliphatic residues in the protein—the higher the value, the higher the thermostability (Ikai, 1980). The aliphatic index of the annotated ENH proteins varied from 9.84 to 109.31. The average GRAVY (Grand Average of Hydropathicity) for the ENH proteins was -0.45 with maximum 0.587 and minimum -1.738 . The GRAVY is calculated based on the total hydropathy values of amino acids divided by the length of the protein. The calculated physicochemical properties of the annotated ENH proteins are provided in **Supplementary File 7**. All the calculated physicochemical properties of the annotated proteins could be useful for further experimental studies of these proteins.

A homology search against the core set of virulent proteins from VFDB resulted in the identification of two virulence proteins—WP_004222859.1 and WP_002890061.1—with comparable similarity to proteins from *Listeria monocytogenes* and *E. coli*, respectively. WP_004222859.1 was identified to be homologous to iap/cwhA which encodes P60 protein, a major extracellular virulence protein in *L. monocytogenes*. This protein has been reported to be involved in pathogen survival and host invasion (Cabanes et al., 2002). From the function prediction, it was found that this protein has an endopeptidase NlpC/P60

domain. The protein WP_002890061.1 was found to be similar to yagV/EcpE, a part of *E. coli* common pilus (ECP). ECP is an extracellular adhesive fiber found in both pathogenic and commensal strains, and is involved in biofilm formation and host cell recognition (Rendón et al., 2007). Targeting virulence factors of a pathogen would hinder the progress of pathogenesis.

VICMPred categorizes proteins into different classes: virulence factors, information molecule, cellular process, and metabolism molecule. It was observed from the prediction that, out of 17 annotated ENH proteins, seven proteins were involved in cellular process, eight were involved in metabolism, and two proteins were involved in information and storage. Furthermore, the MP3 server identified eight proteins as pathogenic and the remaining nine as non-pathogenic proteins. Two proteins (WP_004222859.1 and WP_002890061.1), which were identified as pathogenic by VFDB and MP3, were considered as virulence factors in the present study. The comprehensive results of the predictions are given in Table 2.

Druggability and Subcellular Localization Analysis

Druggability prediction of ENH proteins using DrugBank and ChEMBL revealed that the proteins were not similar to the available known targets. Thus, they can be considered as “novel targets” which can be validated further experimentally.

Determining protein subcellular localization is vital for understanding the role of proteins in a cell, and it also aids in the process of drug discovery and delivery. The current study utilized the subcellular localization tool CELLO, which has high accuracy in predicting subcellular localization of proteins in Gram-negative bacteria. The tool CELLO predicted 10 proteins as cytoplasmic, four as periplasmic, two as inner

membrane proteins, and one as an extracellular protein (Table 2). The proteins predicted in the cytoplasm, periplasm, and inner membrane can be considered as drug targets, and extracellular proteins can be considered as vaccine targets (Barh et al., 2011). Thus, in total, 16 drug targets and one vaccine candidate were identified from the analysis, and this can be further validated using an experimental study.

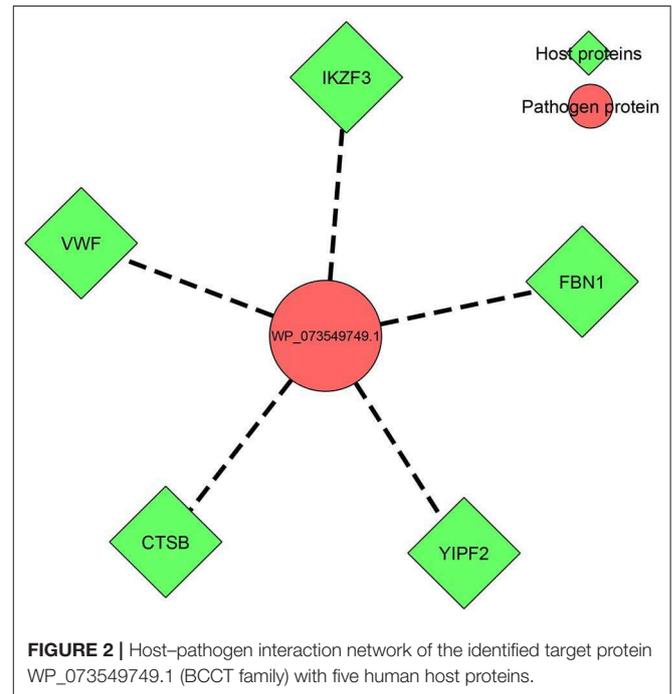


TABLE 2 | List of annotated essential non-homologous proteins with their subcellular localization and virulence prediction.

SI. No	Protein ID	Localization		Virulence prediction		
		CELLO	VICMPred	MP3	VFDB*	
1	WP_002888808.1	Periplasmic	Metabolism molecule	Pathogenic	–	
2	WP_004222859.1	Periplasmic	Metabolism molecule	Pathogenic	Virulent	
3	WP_002890061.1	Periplasmic	Cellular process	Pathogenic	Virulent	
4	WP_002890284.1	Cytoplasmic	Metabolism molecule	Pathogenic	–	
5	WP_004151327.1	Inner membrane	Cellular process	Pathogenic	–	
6	WP_012068456.1	Periplasmic	Information and storage	Pathogenic	–	
7	WP_004176857.1	Cytoplasmic	Information and storage	Pathogenic	–	
8	WP_023288894.1	Cytoplasmic	Cellular process	Non-pathogenic	–	
9	WP_073549749.1	Inner membrane	Metabolism molecule	Pathogenic	–	
10	WP_002898708.1	Cytoplasmic	Cellular process	Non-pathogenic	–	
11	WP_004150795.1	Extracellular	Metabolism molecule	Non-pathogenic	–	
12	WP_041937616.1	Cytoplasmic	Cellular process	Non-pathogenic	–	
13	WP_004143718.1	Cytoplasmic	Metabolism molecule	Non-pathogenic	–	
14	WP_002911528.1	Cytoplasmic	Cellular process	Non-pathogenic	–	
15	WP_002914983.1	Cytoplasmic	Cellular process	Non-pathogenic	–	
16	WP_002918629.1	Cytoplasmic	Metabolism molecule	Non-pathogenic	–	
17	WP_015959101.1	Cytoplasmic	Metabolism molecule	Non-pathogenic	–	

* “–” indicates no significant hits found in VFDB.

Host-Pathogen Interactions

Prediction of host-pathogen interactions among the hypothetical/annotated proteins in the organism could shed insight into understanding the pathogen biology, i.e., the pathogenesis as well as the host response, during

infection/invasion. In the present study, from the 17 ENH proteins, the protein WP_073549749.1, which was annotated as a BCCT family transporter protein, was found to have interactions with five human host proteins, namely, von Willebrand factor, fibrillin-1, protein YIPF2, zinc finger protein Aiolos, and

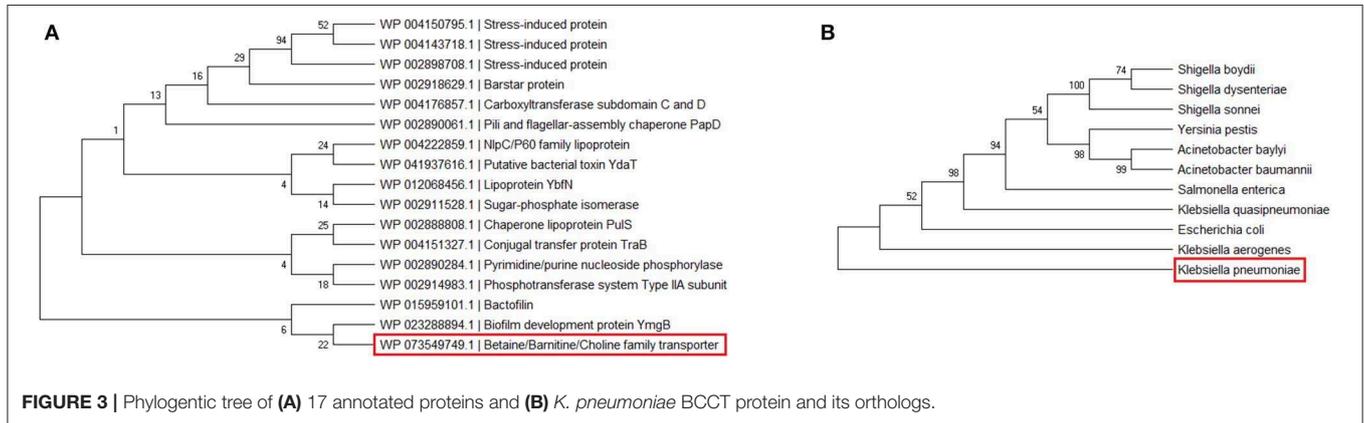


TABLE 3 | List of identified 3D homolog structures of annotated ENH proteins.

Sl. No	Protein ID	PSI-BLAST hit	Organism(s)	Identity (%)	Query coverage (%)	E value	Score	PDB ID
1	WP_002888808.1	-	-	-	-	-	-	-
2	WP_004222859.1	Chain A, Lipoprotein spr	<i>Escherichia coli</i> K-12	43.44	64	2.00E-31	111	2K1G_A
3	WP_002890061.1	Chain A, Probable Fimbrial Chaperone Ecgp	<i>Escherichia coli</i>	28.16	84	4.00E-13	67	5DFK_A
4	WP_002890284.1	Chain A, Upf0345 Protein Vpa0057	<i>Vibrio parahaemolyticus</i> RIMD 2210633	53.76	98	8.00E-28	97.4	2OYZ_A
5	WP_004151327.1	-	-	-	-	-	-	-
6	WP_012068456.1	-	-	-	-	-	-	-
7	WP_004176857.1	Chain B, Ybgj	<i>Escherichia coli</i> K-12	79.36	100	6.00E-126	355	5DUD_B
8	WP_023288894.1	-	-	-	-	-	-	-
9	WP_073549749.1	Chain A, Crystal Structure Of Carnitine Transporter	<i>Escherichia coli</i> K-12	26.67	89	7.00E-38	141	3HFX_A
10	WP_002898708.1	-	-	-	-	-	-	-
11	WP_004150795.1	-	-	-	-	-	-	-
12	WP_041937616.1	Chain A, Crystal Structure Of An Uncharacterized Protein Encoded By Cryptic Prophage	<i>Escherichia coli</i> O6	30.23	44	0.002	37.7	3C4R_A
13	WP_004143718.1	-	-	-	-	-	-	-
14	WP_002911528.1	Chain A, Crystal Structure Of Ribose-5-Phosphate Isomerase Lacab_rpb From <i>Vibrio Parahaemolyticus</i>	<i>Vibrio parahaemolyticus</i> RIMD 2210633	69.52	99	1.00E-104	301	3ONO_A
15	WP_002914983.1	Chain A, PTS system, N, N'-DiacetylchitobiosE-specific IIA component	<i>Escherichia coli</i>	64.58	92	1.00E-36	120	1WCR_A
16	WP_002918629.1	Chain A, Putative Cytoplasmic Protein	<i>Salmonella enterica</i> subsp. enterica serovar Typhimurium str. LT2	70	98	1.00E-42	134	5F4C_A
17	WP_015959101.1	Chain A, Bactofilin A	<i>Caulobacter vibrioides</i> CB15	24.26	60	2.00E-06	47.4	2N3D_A

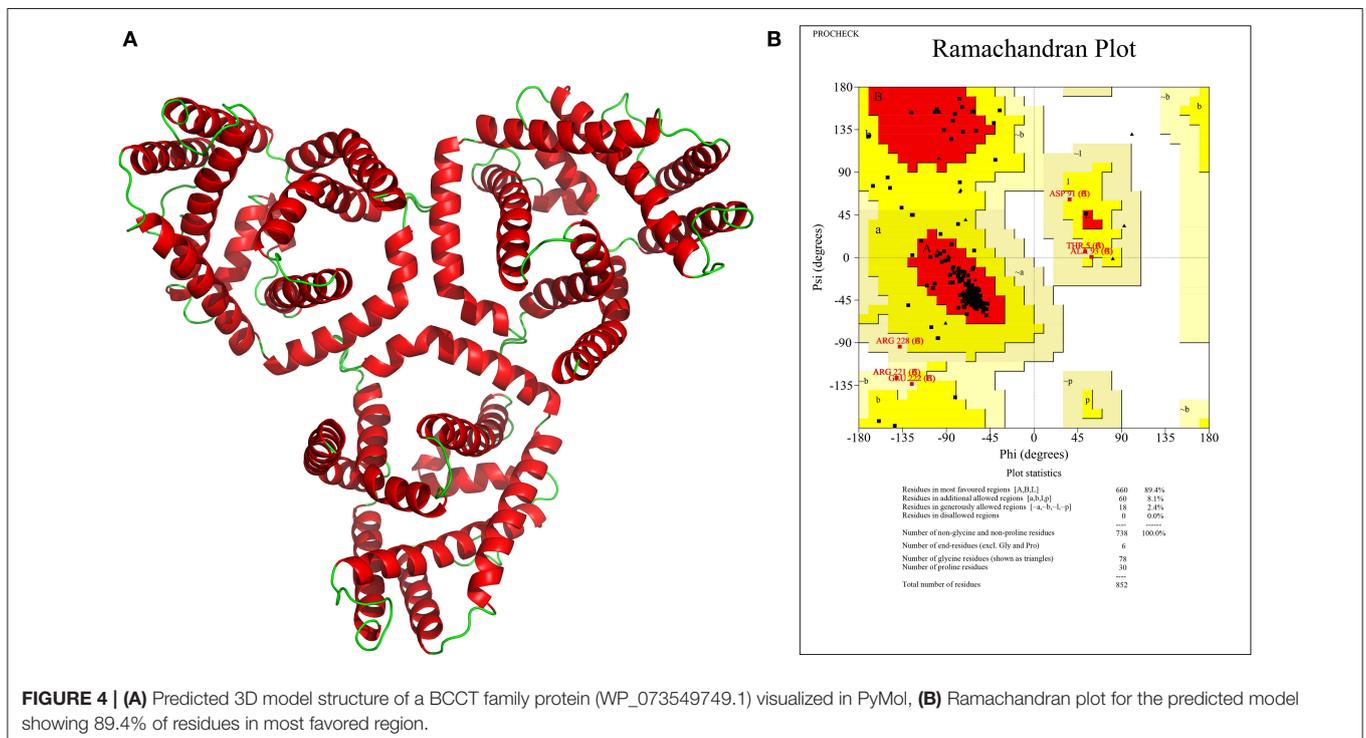
cathepsin B, as depicted in **Figure 2**. This prediction was based on an interlog method in which proteins similar to orthologs in interactions have an increased possibility of interacting with the same partner. The *K. pneumoniae* BCCT protein was found to be similar (75.6% sequence identity) to the putative quaternary ammonium transport protein (UniProt ID: Q8D0R9) from *Yersinia pestis*. The putative quaternary ammonium transport protein is encoded by beT2 gene, which reportedly has shown interactions with the abovementioned five human proteins (**Supplementary File 8**). The functions of the human proteins ranged from binding, catalytic activity, and transcription regulation to biological adhesion.

Phylogenetic analysis of 17 annotated proteins was performed to find out interrelationship among these proteins. The analysis indicated that the BCCT protein, predicted to be involved in host–pathogen interaction, was grouped with biofilm-forming proteins, although they belong to diverse functional classes (**Figure 3A**). However, further study is needed to evaluate the relationship between these two proteins. Moreover, the orthologs of the *K. pneumoniae* BCCT protein were collected from UniProt using a BLASTp search, and a phylogenetic analysis was carried out for those BCCT proteins. It was observed that the *K. pneumoniae* BCCT protein was found to be closely related to *Klebsiella aerogenes* and *E. coli* (**Figure 3B**). Herein, the phylogenetic tree was built using maximum likelihood method with 500 bootstrap replicates in MEGA X, a molecular analysis tool for constructing phylogenetic tree (Kumar et al., 2018). Additionally, a phylogenetic tree was generated using PATRIC for 1,018 genomes belonging to the order Enterobacterales to have an insight into the evolutionary relationship of *K. pneumoniae* with the other members of *Enterobacteriaceae* family

(**Supplementary File 9**). PATRIC is a bioinformatics resource platform that provides multi-omics data and analysis tools for biomedical research (Davis et al., 2020).

Structure Prediction and Validation

Protein structure determination helps in understanding the functional domains responsible for its activity, which is invaluable for novel inhibitor development. The sequences with high similarity tend to have high structural similarity also; a similarity search of the identified 17 drug/vaccine targets against 3D structures of proteins from PDB was thus performed, which identified 10 homologous structures from *E. coli*, *S. enterica*, and *V. parahaemolyticus*. Sequence identity varied from 24.26 to 79.36% with seven proteins showing >30% identity (**Table 3**). Homologous structures can be used as a template for building 3D models for these proteins, which is underway in our laboratory. Here, however, we have reported the structure of an important host–pathogen interacting protein (WP_073549749.1) that showed interactions with the human host. Due to low structural similarity with the available 3D template, the protein structure was modeled using I-TASSER server. Among the five generated models, the best model was chosen based on the C-score (C-score = 1.24), which represents the confidence of the predicted model. The estimated TM-score and RMSD were 0.88 ± 0.07 and 3.8 ± 2.6 Å, respectively. The predicted 3D model of the BCCT protein was trimeric in nature comprising large helical structures. The model was validated using the SAVES server. It was observed that 89.4% of residues of the predicted structure were falling in the most favored regions in the Ramachandran plot while 8.1 and 2.1% were in the additionally and generously allowed region, respectively (**Figure 4**). This emphasized the validity of



our model, and, additionally, molecular dynamics studies can be carried out using this model.

CONCLUSION

Understanding the functions of hypothetical proteins is important since it facilitates the further comprehension of their role in biochemical/physiological pathways and the identification of novel classes of therapeutic targets. The present study utilized an *in silico* approach to identify drug and vaccine targets from hypothetical proteins of *K. pneumoniae*. The study first predicted 30 pathogen-specific essential proteins, for which a functional analysis was carried out. The methodology utilized herein was enabled the annotation of the functions of hypothetical proteins with high confidence. It was found that the proteins have various functional roles as enzymes, lipoproteins, stress-induced proteins, and virulent proteins. However, the functional annotation for some of the proteins was not possible owing to insufficient information. Subcellular localization analysis revealed 16 proteins as drug targets (cytoplasmic, inner membrane, and periplasmic proteins) and one extracellular protein as a vaccine candidate. Structure prediction of one protein which was predicted to be involved in host–pathogen interaction is reported in this study and can be utilized for further experimental studies in this area. In addition, the structural analyses of identified target proteins and screening of potential inhibitors are underway in our laboratory.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

GP and AP conceived and designed the study. GP performed the experiment and analyzed the results. GP, AP, JP, and AR critically

reviewed the analysis and contributed to the preparation of the final version of the manuscript.

ACKNOWLEDGMENTS

GP is grateful to the Department of Biotechnology (DBT), Government of India, for providing the DBT-BINC Junior Research Fellowship. JP is thankful to the Indian Council of Medical Research (ICMR), Government of India, for the ICMR Senior Research Fellowship. AR is grateful to Pondicherry University for the pre-doctoral fellowship. Authors are indebted to Centre for Bioinformatics, Pondicherry University, Pondicherry, for providing the computational facility to carry out the work. DBT, DIT, DST-FIST, and UGC-SAP, Government of India support research works carried out at Centre for Bioinformatics, Pondicherry University, Pondicherry.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2020.00109/full#supplementary-material>

Supplementary File 1 | List of all hypothetical proteins from *Klebsiella pneumoniae* considered in the study.

Supplementary File 2 | List of shortlisted essential non-homologous proteins.

Supplementary File 3 | List of identified uncharacterized domain family essential non-homologous (ENH) proteins.

Supplementary File 4 | Identified essential proteins and the respective homologs from DEG.

Supplementary File 5 | Identified essential non-homologous proteins along with predicted GO using CELLO2GO and GO_FEAT.

Supplementary File 6 | Mapped ENH proteins on STRING PPI data.

Supplementary File 7 | Calculated physicochemical properties of the identified essential non-homologous proteins.

Supplementary File 8 | List of host proteins interacting with a BCCT family protein.

Supplementary File 9 | Phylogenetic tree of 1,018 genomes from the order Enterobacterales generated using PATRIC.

REFERENCES

- Altschul, S. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Araujo, F. A., Barh, D., Silva, A., Guimarães, L., and Ramos, R. T. J. (2018). GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. *Sci. Rep.* 8:1794. doi: 10.1038/s41598-018-20211-9
- Ashurst, J. V., and Dawson, A. (2019). “*Klebsiella pneumoniae*” in *StatPearls* (Treasure Island, FL: StatPearls Publishing). Available online at: <https://www.ncbi.nlm.nih.gov/pubmed/30085546>
- Baraniak, A., Izdebski, R., Zabicka, D., Bojarska, K., Górska, S., Literacka, E., et al. (2017). Multiregional dissemination of KPC-producing *Klebsiella pneumoniae* ST258/ST512 genotypes in Poland, 2010–14. *J. Antimicrob. Chemother.* 72, 1610–1616. doi: 10.1093/jac/dkx054
- Barh, D., Tiwari, S., Jain, N., Ali, A., Santos, A. R., Misra, A. N., et al. (2011). *In silico* subtractive genomics for target identification in human bacterial pathogens. *Drug Dev. Res.* 72, 162–177. doi: 10.1002/ddr.20413
- Barnhart, M. M., Pinkner, J. S., Soto, G. E., Sauer, F. G., Langermann, S., Waksman, G., et al. (2000). PapD-like chaperones provide the missing information for folding of pilin proteins. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7709–7714. doi: 10.1073/pnas.130183897
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235
- Buckle, A. M., Schreiber, G., and Fersht, A. R. (1994). Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution. *Biochemistry* 33, 8878–8889. doi: 10.1021/bi00196a004
- Bzowska, A. (2015). Purine and pyrimidine nucleoside phosphorylases—remarkable enzymes still not fully understood. *Postepy Biochem.* 61, 260–273.
- Cabanes, D., Dehoux, P., Dussurget, O., Frangeul, L., and Cossart, P. (2002). Surface proteins and the pathogenic potential of *Listeria monocytogenes*. *Trends Microbiol.* 10, 238–245. doi: 10.1016/S0966-842X(02)02342-9
- CDC.gov. (2020). *Klebsiella pneumoniae* in *Healthcare Settings*. HAI; CDC. Available online at: <https://www.cdc.gov/hai/organisms/klebsiella/klebsiella.html> (accessed February 14, 2020).
- Chandler, J. R., and Dunny, G. M. (2004). Enterococcal peptide sex pheromones: synthesis and control of biological activity. *Peptides* 25, 1377–1388. doi: 10.1016/j.peptides.2003.10.020

- Chen, L., Zheng, D., Liu, B., Yang, J., and Jin, Q. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* 44, D694–D697. doi: 10.1093/nar/gkv1239
- Chen, Y. C., Lu, C. H., and Hwang, J. K. (2006). Prediction of protein subcellular localization. *Proteins* 64, 643–651. doi: 10.1002/prot.21018
- Choi, S. B., Normi, Y. M., and Wahab, H. A. (2009). Why hypothetical protein KPN00728 of *Klebsiella pneumoniae* should be classified as chain C of succinate dehydrogenase? *Protein J.* 28, 415–427. doi: 10.1007/s10930-009-19209-9
- Chung, P. Y. (2016). The emerging problems of *Klebsiella pneumoniae* infections: carbapenem resistance and biofilm formation. *FEMS Microbiol. Lett.* 363:fnw219. doi: 10.1093/femsle/fnw219
- Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., et al. (2020). The PATRIC bioinformatics resource center: expanding data and analysis capabilities. *Nucleic Acids Res.* 48, D606–D612. doi: 10.1093/nar/gkz943
- de Campos, P. A., Royer, S., Batistão, D. W., da F., Araújo, B. F., Queiroz, L. L., et al. (2016). Multidrug resistance related to biofilm formation in acinetobacter baumannii and *Klebsiella pneumoniae* clinical strains from different pulsotypes. *Curr. Microbiol.* 72, 617–627. doi: 10.1007/s00284-016-0996-x
- Duval, V., Foster, K., Brewster, J., and Levy, S. B. (2017). A novel regulatory cascade involving BluR, YcgZ, and lon controls the expression of OmpF porin. *Front. Microbiol.* 8:1148. doi: 10.3389/fmicb.2017.01148
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). "Protein identification and analysis tools on the ExPASy server," in *The Proteomics Protocols Handbook*, ed J. M. Walker (Totowa, NJ: Humana Press Inc.), 571–607. doi: 10.1385/1-59259-890-0:571
- Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., et al. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1100–D1107. doi: 10.1093/nar/gkr777
- Gill, S. C., and von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182, 319–326. doi: 10.1016/0003-2697(89)90602-7
- Goodacre, N. F., Gerloff, D. L., and Uetz, P. (2014). Protein domains of unknown function are essential in bacteria. *mBio* 5:e00744-13. doi: 10.1128/mBio.00744-13
- Gupta, A., Kapil, R., Dhakan, D. B., and Sharma, V. K. (2014). MP3: a software tool for the prediction of pathogenic proteins in genomic and metagenomic data. *PLoS ONE* 9:e93907. doi: 10.1371/journal.pone.0093907
- Hartley, R. W. (1988). Barnase and barstar. Expression of its cloned inhibitor permits expression of a cloned ribonuclease. *J. Mol. Biol.* 202, 913–915. doi: 10.1016/0022-2836(88)90568-2
- Hassan, M. I., Ahmad, F., Ansari, A., Islam, A., Rasool, M., Anjum, F., et al. (2016). Functional annotation of putative conserved proteins from *Borrelia burgdorferi* to find potential drug targets. *Int. J. Comput. Biol. Drug Des.* 9:295. doi: 10.1504/IJCBDD.2016.10000929
- Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *J. Biochem.* 88, 1895–1898.
- Islam, M. S., Shahik, S. M., Soheli, M., Patwary, N. I. A., and Hasan, M. A. (2015). *In silico* structural and functional annotation of hypothetical proteins of vibrio cholerae O139. *Genomics Inform.* 13, 53–59. doi: 10.5808/GI.2015.13.2.53
- Jadhav, A., Shanmugham, B., Rajendiran, A., and Pan, A. (2014). Unraveling novel broad-spectrum antibacterial targets in food and waterborne pathogens using comparative genomics and protein interaction network analysis. *Infect. Genet. Evol.* 27, 300–308. doi: 10.1016/j.meegid.2014.08.007
- Kaye, K. S., and Dhar, S. (2016). Infection prevention and control in healthcare, part ii: epidemiology and prevention of infections. *Infect. Dis. Clin. North. Am.* 30, xiii–xiv. doi: 10.1016/j.idc.2016.09.001
- Krapp, F., Ozer, E. A., Qi, C., and Hauser, A. R. (2018). Case report of an extensively drug-resistant *Klebsiella pneumoniae* infection with genomic characterization of the strain and review of similar cases in the United States. *Open Forum Infect. Dis.* 5:ofy074. doi: 10.1093/ofid/ofy074
- Kühn, J., Briegel, A., Mörschel, E., Kahnt, J., Leser, K., Wick, S., et al. (2010). Bactofilins, a ubiquitous class of cytoskeletal proteins mediating polar localization of a cell wall synthase in *Caulobacter crescentus*. *EMBO J.* 29, 327–339. doi: 10.1038/emboj.2009.358
- Kumar, R., and Nanduri, B. (2010). HPIDB—a unified resource for host-pathogen interactions. *BMC Bioinformatics* 11:S16. doi: 10.1186/1471-2105-11-S6-S16
- Kumar, S., Stecher, G., Li, M., Nknyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Lee, H. S., Lee, S., Kim, J.-S., Lee, H.-R., Shin, H.-C., Lee, M.-S., et al. (2019). Structural and physiological exploration of *Salmonella* Typhi YfdX uncovers its dual function in bacterial antibiotic stress and virulence. *Front. Microbiol.* 9:3329. doi: 10.3389/fmicb.2018.03329
- Liu, C., and Guo, J. (2019). Hypervirulent *Klebsiella pneumoniae* (hypermucoviscous and aerobactin positive) infection over 6 years in the elderly in China: antimicrobial resistance patterns, molecular epidemiology and risk factor. *Ann. Clin. Microbiol. Antimicrob.* 18:4. doi: 10.1186/s12941-018-0302-9
- Luo, H., Lin, Y., Gao, F., Zhang, C.-T., and 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–D580. doi: 10.1093/nar/gkt1131
- Masuda, N., and Church, G. M. (2002). *Escherichia coli* gene expression responsive to levels of the response regulator EvgA. *J. Bacteriol.* 184, 6225–6234. doi: 10.1128/JB.184.22.6225-6234.2002
- McClelland, M., Sanderson, K. E., Spieth, J., Clifton, S. W., Latreille, P., Courtney, L., et al. (2001). Complete genome sequence of *Salmonella enterica* serovar typhimurium LT2. *Nature* 413, 852–856. doi: 10.1038/35101614
- Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47, D351–D360. doi: 10.1093/nar/gky1100
- Naqvi, A. A. T., Anjum, F., Khan, F. I., Islam, A., Ahmad, F., and Hassan, M. I. (2016). Sequence analysis of hypothetical proteins from *Helicobacter pylori* 26695 to identify potential virulence factors. *Genomics Inform.* 14, 125–135. doi: 10.5808/GI.2016.14.3.125
- Ocampo, A. M., Chen, L., Cienfuegos, A. V., Roncancio, G., Chavda, K. D., Kreiswirth, B. N., et al. (2016). A two-year surveillance in five colombian tertiary care hospitals reveals high frequency of non-CG258 clones of carbapenem-resistant *Klebsiella pneumoniae* with distinct clinical characteristics. *Antimicrob. Agents Chemother.* 60, 332–342. doi: 10.1128/AAC.01775-15
- Postma, P. W., Lengeler, J. W., and Jacobson, G. R. (1993). Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* 57, 543–594. doi: 10.1128/MMBR.57.3.543-594.1993
- Prava, J., G. P., and Pan, A. (2018). Functional assignment for essential hypothetical proteins of *Staphylococcus aureus* N315. *Int. J. Biol. Macromol.* 108, 765–774. doi: 10.1016/j.ijbiomac.2017.10.169
- Rendón, M. A., Saldaña, Z., Erdem, A. L., Monteiro-Neto, V., Vázquez, A., Kaper, J. B., et al. (2007). Commensal and pathogenic *Escherichia coli* use a common pilus adherence factor for epithelial cell colonization. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10637–10642. doi: 10.1073/pnas.0704104104
- Riquelme, S. A., Ahn, D., and Prince, A. (2018). *Pseudomonas aeruginosa* and *Klebsiella pneumoniae* adaptation to innate immune clearance mechanisms in the lung. *J. Innate Immun.* 10, 442–454. doi: 10.1159/000487515
- Robbe-Saule, V., Lopes, M. D., Kolb, A., and Norel, F. (2007). Physiological effects of Crl in *Salmonella* are modulated by sigmaS level and promoter specificity. *J. Bacteriol.* 189, 2976–2987. doi: 10.1128/JB.01919-06
- Saha, S., and Raghava, G. P. S. (2006). VICMPred: an SVM-based method for the prediction of functional proteins of Gram-negative bacteria using amino acid patterns and composition. *Genomics Proteomics Bioinformatics* 4, 42–47. doi: 10.1016/S1672-0229(06)60015-6
- Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., et al. (2019). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 47, D23–D28. doi: 10.1093/nar/gky1069
- Shahbaaz, M., Bisetty, K., Ahmad, F., and Hassan, M. I. (2016). Current advances in the identification and characterization of putative drug and vaccine targets in the bacterial genomes. *Curr. Top. Med. Chem.* 16, 1040–1069. doi: 10.2174/1568026615666150825143307
- Shahbaaz, M., Hassan, M. I., and Ahmad, F. (2013). Functional annotation of conserved hypothetical proteins from *Haemophilus influenzae* Rd KW20. *PLoS ONE* 8:e84263. doi: 10.1371/journal.pone.0084263

- Soo, V. W. C., Hanson-Manful, P., and Patrick, W. M. (2011). Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1484–1489. doi: 10.1073/pnas.1012108108
- Sotgiu, G., Are, B. M., Pesapane, L., Palmieri, A., Muresu, N., Cossu, A., et al. (2018). Nosocomial transmission of carbapenem-resistant *Klebsiella pneumoniae* in an Italian university hospital: a molecular epidemiological study. *J. Hosp. Infect.* 99, 413–418. doi: 10.1016/j.jhin.2018.03.033
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131
- Takahashi, N., Hayano, T., and Suzuki, M. (1989). Peptidyl-prolyl cis-trans isomerase is the cyclosporin A-binding protein cyclophilin. *Nature* 337, 473–475. doi: 10.1038/337473a0
- Teh, B. A., Choi, S. B., Musa, N., Ling, F. L., Cun, S. T. W., Salleh, A. B., et al. (2014). Structure to function prediction of hypothetical protein KPN_00953 (Ycbk) from *Klebsiella pneumoniae* MGH 78578 highlights possible role in cell wall metabolism. *BMC Struct. Biol.* 14:7. doi: 10.1186/1472-6807-14-7
- Turab Naqvi, A. A., Rahman, S., Rubi Zeya, F., Kumar, K., Choudhary, H., et al. (2017). Genome analysis of *Chlamydia trachomatis* for functional characterization of hypothetical proteins to discover novel drug targets. *Int. J. Biol. Macromol.* 96, 234–240. doi: 10.1016/j.ijbiomac.2016.12.045
- Vuotto, C., Longo, F., Balice, M. P., Donelli, G., and Varaldo, P. E. (2014). Antibiotic resistance related to biofilm formation in *Klebsiella pneumoniae*. *Pathogens* 3, 743–758. doi: 10.3390/pathogens3030743
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi: 10.1093/nar/gkx1037
- Wong, M. T., Choi, S. B., Kuan, C. S., Chua, S. L., Chang, C. H., Normi, Y. M., et al. (2012). Structural modeling and biochemical characterization of recombinant KPN_02809, a zinc-dependent metalloprotease from *Klebsiella pneumoniae* MGH 78578. *Int. J. Mol. Sci.* 13, 901–917. doi: 10.3390/ijms13010901
- Xu, L., Sun, X., and Ma, X. (2017). Systematic review and meta-analysis of mortality of patients infected with carbapenem-resistant *Klebsiella pneumoniae*. *Ann. Clin. Microbiol. Antimicrob.* 16:18. doi: 10.1186/s12941-017-0191-3
- Xu, Q., Sudek, S., McMullan, D., Miller, M. D., Geierstanger, B., Jones, D. H., et al. (2009). Structural basis of murein peptide specificity of a gamma-D-glutamyl-l-diamino acid endopeptidase. *Structure* 17, 303–313. doi: 10.1016/j.str.2008.12.008
- Yamaguchi, Y., and Inouye, M. (2011). Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nat. Rev. Microbiol.* 9, 779–790. doi: 10.1038/nrmicro2651
- Yang, J., and Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* 43, W174–W181. doi: 10.1093/nar/gkv342
- Yang, Z., Zeng, X., and Tsui, S. K.-W. (2019). Investigating function roles of hypothetical proteins encoded by the *Mycobacterium tuberculosis* H37Rv genome. *BMC Genomics* 20:394. doi: 10.1186/s12864-019-5746-6
- Yu, C.-S., Cheng, C.-W., Su, W.-C., Chang, K.-C., Huang, S.-W., Hwang, J.-K., et al. (2014). CELLO2GO: a web server for protein subcellular localization prediction with functional gene ontology annotation. *PLoS ONE* 9:e99368. doi: 10.1371/journal.pone.0099368
- Yu, H., Luscombe, N. M., Lu, H. X., Zhu, X., Xia, Y., Han, J.-D. J., et al. (2004). Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res.* 14, 1107–1118. doi: 10.1101/gr.1774904
- Zhang, R.-G., Andersson, C. E., Skarina, T., Evdokimova, E., Edwards, A. M., Joachimiak, A., et al. (2003b). The 2.2 Å resolution structure of RpiB/AlsB from *Escherichia coli* illustrates a new approach to the ribose-5-phosphate isomerase reaction. *J. Mol. Biol.* 332, 1083–1094. doi: 10.1016/j.jmb.2003.08.009
- Zhang, R. G., Andersson, C. E., Savchenko, A., Skarina, T., Evdokimova, E., Beasley, S., et al. (2003a). Structure of *Escherichia coli* ribose-5-phosphate isomerase: a ubiquitous enzyme of the pentose phosphate pathway and the Calvin cycle. *Structure* 11, 31–42. doi: 10.1016/S0969-2126(02)00933-4
- Ziegler, C., Bremer, E., and Krämer, R. (2010). The BCCT family of carriers: from physiology to crystal structure. *Mol. Microbiol.* 78, 13–34. doi: 10.1111/j.1365-2958.2010.07332.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Pranavathiyani, Prava, Rajeev and Pan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.