Int. J. Pharm. Investigation, 2021;11(1):32-40

Original Article

Important Insight of Hypothetical Proteins from *E. faecium* Strain 13-022, the Drug Targeted Virulent Big-1 Adhesins: An *in-silico* Approach

Brindangnanam Pownraj[1,*] Meenatchi Ramu[2], Ekramul Haque[2]
[1]Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry, INDIA.
[2]Department of Microbiology, School of Life Sciences, Pondicherry University, Puducherry, INDIA.

ABSTRACT

Background: *Enterococcus faecium* is an emerging multidrug resistant opportunistic pathogen responsible for causing most of the nosocomial infections. Adhesins are the cell wall anchoring proteins implicated in the pathogenesis of enterococcal infections. The present investigation is carried out to spot the occurrence of bacterial immunoglobulin-like (Ig) domain-1 or Big-1 adhesins among 204 HPs from *E. faecium* strain 13-022. **Methods:** The pathogen *E. faecium* chromosomal strain 13-022 was searched in the NCBI database which comprised of 2746 proteins. To filter HPs, the keyword 'hypothetical proteins' with sequence length >100 residues was queried. The filtered 204 HPs were subjected to functional annotation, physico-chemical properties, virulence factors, cellular location, secondary structure prediction and protein-protein interactions (PPIs). Finally, 3D models were obtained for essential non-homologous adhesins with potential drug binding pockets. **Results:** Primarily functional classification of 204 HPs which are assigned to 27 different functional activities with good thermal stability (50%), hydropathicity and virulence factors (79%). Majority of the HPs are predicted to reside in the cytoplasm and cell membrane. 78 HPs are predicted with high confidence. Among them, 14 are having βαβ motifs including two adhesins and the PPI network has 4 gene set of Mga helix-turn-helix and 2 gene set of putative adhesion and 77 proteins are essential hypothetical proteins (EHPs). Of 77 EHPs, 65 are pathogen-specific, indeed considered as probable drug targets. In these 65 essential pathogen specific proteins, 23 targets are found to be biological targets and rest are novel targets. Among 23 targets, three are adhesins those have therapeutic applications. **Conclusion:** The present study predicted the occurrence of virulent drug targeted Big-1 specifically bacterial non-pilus fimbriae immunoglobulin-like (Ig) domain-1among 204 HPs. Its structure, function and significance were emphasized to develop novel drugs for better treatment.

Keywords: *Enterococcus faecium*, Big-1 adhesins, βαβ motifs and Drug targets, Protein-protein Interactions, Essential Hps.

Correspondence

Ms. Brindangnanam Pownraj

Research Scholar, Center for Bioinformatics, Pondicherry University, Kalapet, Pondicherry-605014, INDIA.

Phone: +91 9444589478

Email: astra.bioinfo15@gmail.com

DOI: 10.5530/ijpi.2021.1.7

INTRODUCTION

Enterococcus faecium is a Gram-positive Lactobacillus which belongs to the Phylum Firmicutes. It is usually present in the gastrointestinal tract of animals, soil, water, food, hospital outbreaks and surgical equipments.[1] According to Gao (2018), its mode of action is found to be sporadically associated opportunistic pathogens.[2] At the same time, 1% of human gut microbiota is constituted by *Enterococcus* thus, it is traditionally believed as a commensal bacterium. For the past five decades, *E. faecium* is emerging as one of the leading infectious bacterium as it is liable for most of the nosocomial infections.[1,3] The Intensive Care Unit (ICUs) 2016 surveillance report states that *E. faecium* is responsible for common healthcare-associated infections such as urinary tract infection, bloodstream infection, pneumonia, surgical site infections and most of the antimicrobial resistant pathogens of *Enterococcus* species.[4]

Several aggregates and virulent factors are involved in pathogenicity to promote *Enterococcus* tenacity: adhesins are one of them. Adhesins are the first entity interacting with the host by means of colonization/ adhesions as a primary cause of infection or helps in stable surface binding. There are two types of adhesins, fimbrial and afimbrial. The classical example of bacterial adhesin structures are fimbriae or pilus. Fimbrial adhesins belong to group of pilus-specific housekeeping sortases.[5] Basically, these are long polymeric structures linked covalently by single chain of protein subunits ended up with two forms (major and minor) of adhesins liable for cell wall anchoring and pathogenicity [Figure 1].[6]

Afimbrial adhesins (non-pilus) are the group of proteins lacking long polymeric fimbrial structures, that generally mediates additional close contact with the host cell which occurs over a shorter range than with

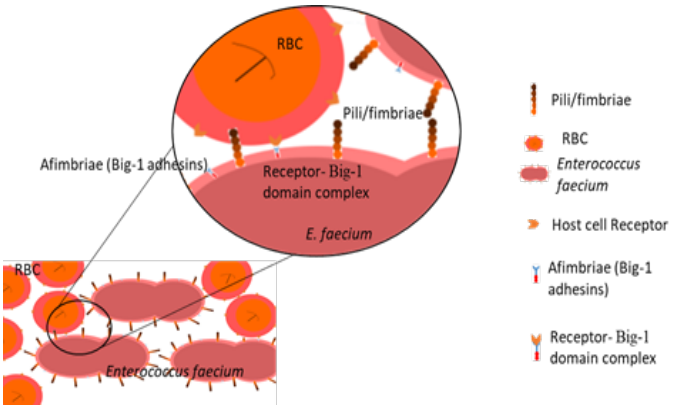


Figure 1: Afimbrial (Big-1 adhesins) adhesins interacting with host red blood cells (RBC).

fimbriae.[7,8] As a virulent factor, they play a key role in the host cell surface colonization.[9] Furthermore, intra and extracellular adhesins promote clonal population of *Enterococcus* thus causes infections.[10,11] Genetic acquisition and evolution of *E. faecium* are the primary reason for the multidrug resistance.[12,13] This suggests the need to develop new drugs for preventing and precluding these infections.

The genes with indefinite homologous termed hypothetical or unknown/uncharacterized because it is unclear whether they encode actual protein function.[14-16] Several bioinformatics tools and soft-wares have been developed to predict the functions of unknown proteins based on the sequences. Since November 11th, 2020, the NCBI database was found to contain 2061 genome assemblies and annotation reports of *E. faecium*, out of which 155 are complete genome sequence data, 17 chromosomal sequences, 866 scaffold sequences and 1023 contig sequences. At chromosome level, the GC content varies between 37-40% with chromosome and plasmid replicons. Each chromosome has 2000 to 3400 proteins, out of these 240 to 780 are hypothetical proteins (HPs).

A large number of HPs from *E. faecium* chromosomal strain 13-022 were annotated in this study. Further, classification of HPs into different functional categories reveal the presence of virulent adhesins and their functions and structures. Studies on HPs suggests the significance of predicting the functions of unknown proteins and identification of novel drug targets for the improving the treatment methods of various nosocomial infections.

MATERIALS AND METHODS

Retrieval of the hypothetical protein

The *E. faecium* chromosomal strain was searched in the database of NCBI using the keyword, "*Enterococcus faecium*". The proteins from genome assembly and annotation report of *E. faecium* strain 13-022 was filtered by hitting the search for "hypothetical protein". With the retrieved HPs, the primary, secondary and tertiary analysis were performed as depicted in the Figure 2.

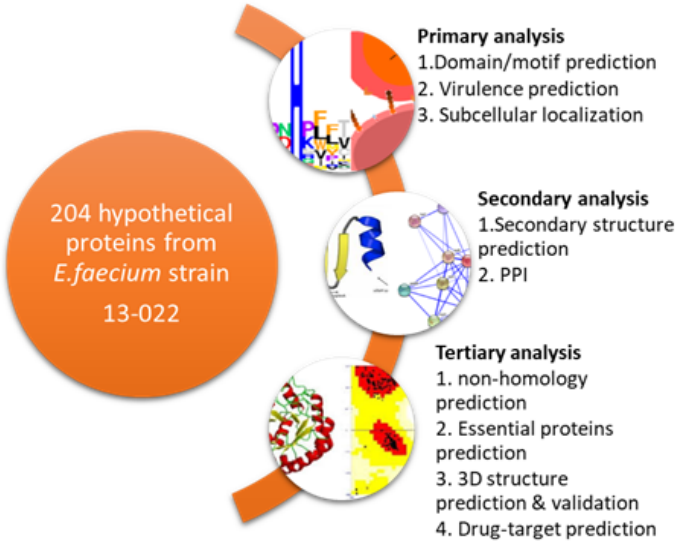


Figure 2: Complete workflow.

Primary analysis

Domain/motif prediction

The publically available bioinformatics tools and databases such as NCBI PSI-BLAST, Conserved Domain Database (CDD), Pfam and InterProScan were used to uncover the functions of HPs. Homologous proteins were predicted by searching sequences with high similarity and identity using BLAST and distantly related proteins were predicted using PSI-BLAST (based on position specific scoring matrix (PSSM)). Functional motifs/domains present in HPs were predicted using CDD,[17] Pfam[18] and InterProScan.[19] Sequences with high confidence were subjected to further analysis.

Physicochemical properties of HPs

The theoretical physicochemical properties of HPs were computed using ProtParam tool available at ExPASy server. The output parameters were isoelectric point (pI), molecular weight, instability index,[20] aliphatic index,[21] and grand average hydropathy (GRAVY).[22]

Subcellular localization prediction

The subcellular localization of HPs was predicted by CELLO2GO and Gpos-mPLoc. CELLO2GO (http://cello.life.nctu.edu.tw/cello2go/) uses SVM (support vector machine) and BLAST-homology searching approaches to find the subcellular localization and Gene Ontology (GO).[23] Gpos-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/Gpos-multi) is one of the package from Cell-PLoc 2.0 server, specific for predicting subcellular localization of gram positive bacterial proteins.[24]

Virulence prediction of HPs

To predict the virulent HPs, VICMpred and Virulentpred were used. VICMpred (http://crdd.osdd.net/raghava/vicmpred/) is a support vector machine (SVM) based web-application server aimed to categorize the given protein into virulence factors, information molecule, cellular process and metabolism molecule.[25] The Virulentpred server (http://bioinfo.icgeb.res.in/virulent/) uses bi-layer cascade SVM based method having domain/motif patterns to predict the bacterial virulent proteins.[26]

Secondary analysis

Complex super-secondary structure prediction

The core complex super secondary structure in protein is loop-helix-loop (βαβ motifs) where functional active sites commonly occur.[27] To predict the βαβ motifs of HPs, consensus secondary structure prediction tool (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_seccons.html) is used. It is based on database of secondary structure assignment for all protein entries (DSSP) with significant parameters (Garnier,[28] PREDATOR,[29] SOPM (a Self-Optimized Method for protein secondary structure prediction)[30] and SOPMA (Self-Optimized Prediction Method with Alignment)[31] to predict the conformational state of each residue.

Prediction of protein-protein interaction

STRING (https://string-db.org/) database was used to predict the protein-protein interactions (PPI). PPIs are derived from five main sources: Genomic context predictions, advanced *in-silico* lab experiments and co-expression, automated text mining and prior knowledge/information.[32] Multiple protein sequences (HPs with high confidence) were searched by specifying the organism name '*Enterococcus faecium*'. The list of matched proteins was used to construct the PPI network.

Tertiary analysis

Essential HPs and Human non-homologous prediction

Essential genes are important for the livelihood of all domains of life. Their product comprises excellent targets for antibacterial drugs.[33] To predict them the query sequences were subjected to BLAST against the Database of Essential Genes (DEG) (http://www.essentialgene.org/), which contains all the essential genes that are currently available.

To predict the human non-homologous proteins, NCBI non-redundant database queried with essential HPs with the following parameters: organism (limited): human, e-value threshold: 0.0001 and a bit score cut-off: 100. Protein sequences displayed considerable match with human proteome were excluded from further analysis.[15,34]

Prediction of novel Drug targets

The sieved HPs from the previous step were subjected to a homology search against the DrugBank and chEMBL (https://www.ebi.ac.uk/chembl) database to confirm the drug gable property of the given protein.[35] In contrast, the lack of hit represents the novel targets.

3D structure prediction and evaluation

The druggable targets specifically adhesins were subjected to I-TASSER sever (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). It generates the sequence based 3D (3-Dimensional) models by predicting secondary structure, solvent accessibility, normalized B-factor, templates, ligand binding sites, EC-number, active sites and GO.[36] Besides, structure evaluation was performed by Ramachandran plot using PROCHECK (https://servicesn.mbi.ucla.edu/PROCHECK/).

RESULTS

The term '*Enterococcus faecium*' was searched in the NCBI database, which bring about 2061 genome assembly and annotation entries. Among them, enterococcus chromosomal strain 13-022 was selected which comprised of 2746 proteins. To acquire HPs, the keyword 'hypothetical proteins' was used and 503 entries were obtained. Further, HPs with sequence length >100 residues were filtered and retrieved and they were around 204 proteins. These filtered HPs were subjected to functional and properties prediction over series of analysis. Primary analysis provides homologous functional motif/domain, physico-chemical properties, subcellular location and virulence factors. In secondary analysis, sequence based secondary structure depicted the potential active sites alongside PPI interactions. In tertiary analysis, 3D models were obtained for essential non-homologous adhesins with potential drug binding pockets.

The sequences of HPs were subjected to predict the motif/domain using various bioinformatics tools as mentioned in the methodology. Out of 204 HPs, 78 HPs were predicted with well-known structural domain/motif. According to Prava (2018) and literature searches, HPs were assigned with 27 different functional activities including adhesins[15] [Supplementary Table 1, Table 1 and Figure 3].

The amino acid sequences of 204 HPs were analyzed to calculate their physico-chemical properties and the result is shown in Supplementary Table 2. The molecular weight (MW) of the all predicted HPs was ranged between 10000 to 108000 Da. The intolerance of surrounding charge of the protein was predicted through isoelectric point (pI) which ranges from 4.0 to 11.0. The thermostability of the HPs was determined by high aliphatic index value ranges from 40.0 to 162.0. The better interaction of HPs with water disclosed by low grade average of hydropathy ranges (GRAVY) from -1.0 to 6.0. The stability of the protein was determined by the value of instability index which is <40. Out of 204 HPs, 118 HPs were stable including adhesins (WP_002286311.1 and WP_002340339) [Supplementary Table 2].

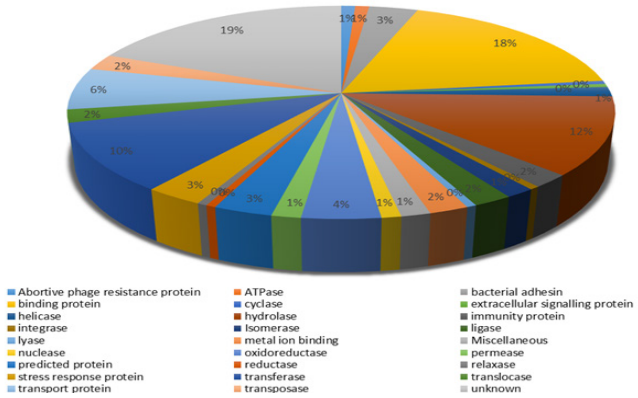


Figure 3: The functional assignment of HPs from *E. faecium* 13-022.

Subcellular localization analysis of HPs helps in classifying them as drug and vaccine targets. Via Gpos-mPLo, 87 membrane proteins, 56 extracellular and 61 biased proteins were localized. In a total of 5 adhesins, 4 were predicted to present in extracellular membrane and 1 in cell membrane. Using CELLO2GO, 124 soluble cytoplasmic proteins, 24 extracellular, 53 membrane proteins, 1 cell wall protein and 1 inner membrane protein were predicted with gene_codes and their functions [Supplementary Table 3]. VICMpred depicted that out of 204 HPs, 107(52.45%) are involved in cellular process, 16 (7.85%) in information storage, 67 (32.84%) in metabolism and 14 (6.86%) act as virulence factors. From, Virulentpred, 79% HPs were predicted to be the virulent and the remaining 21% were non-virulent (Supplementary Table 3).

According to Sun (2016),[27] the active functional sites were predicted through searching for the βαβ motif pattern (eeeccccccchhhhhhhccceeee) in the HP's sequences through consensus secondary structure prediction tool. Among 78, 12 HPs were having this pattern altogether two adhesin (WP_002289495.1 and WP_002345015.1) proteins (Table 1).

Using STRING database, these 78 HPs (Table 1) were subjected to the construction of PPI network. Subsequently, 49 proteins from *E. faecium* C68 and 44 proteins from *E. faecium* NRRLB2354 were the two sets of output matched with the input *E. faecium* 13-022 HPs sequences. The *E. faecium* C68 was selected to construct the PPI network as of maximum hit. The network statistics displayed 57 nodes, 68 edges with *p*-value of < 1.0e-16, 2.39 average node degree and 0.43 average local clustering co-efficient. The functional enrichment of the network has the following pfam domains, 2 gene set of putative adhesion domain and 4 gene set of Mga helix-turn-helix domain with the false discovery rate of 0.0089 and 0.0379, respectively [Figure 4].

Same set of proteins (78 HPs) were subjected to search against DEG database. It showed hit for 77 HPs and considered as the essential HPs (EHPs). These 77 EHPs were subjected to protein BLAST (BLASTp) search against Homo sapiens (taxid: 9606) proteome of the non-redundant database with significant parameters. Out of which, 65 proteins were exclusively present on pathogen and predicted as the potential drug targets. While excluding the significant hits, out of 5 adhesin, 2 showed similar match with human proteome (excluded from the list subsequently) and the residual were considered as novel targets.

These 65 human non-homologous HPs were taken for homology search against DrugBank/chEMBL to find the potential drug targets. Among them, 23 HPs including 3 adhesins were showed 50 to 68.6% similarity with known targets in the database (3-oxoacyl-[acyl-carrier-protein] synthase 2, 70S-ribosome, aminoglycoside acetyltransferase, botulinum neurotoxin-typeF, DNA polymerase-III-polC-type, fructose bisphosphate aldolase class-2, glutamyl endopeptidase, HTH-type-MgrA, outer membrane protein TolC, penicillin-binding protein,

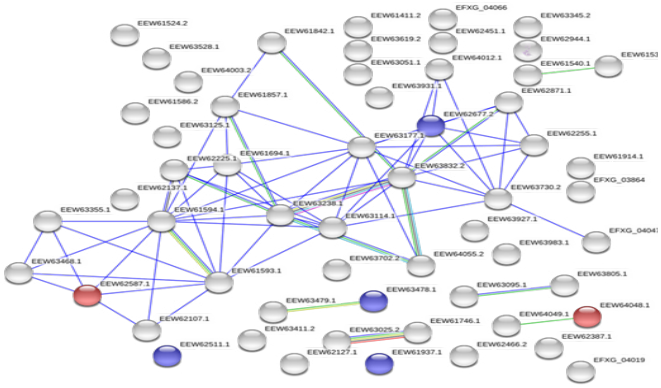


Figure 4: PPI interaction network for the 78 HPs with high confidence showed that EEW63468.1 (LOR-superfamily), EEW63355.1 (baeRF_family6), EEW61594.1, EEW61593.1 (YycH protein) and EEW62107.1 (Halogen_Hydrol) were interacting with EEW62587.1 (putative_adhesin (gene set-1-red sphere)) at left side corner and EEW64049.1 (DUF5348) solely interacting with EEW64048.1 (Accession_no. WP_002286311.1-putative_adhesin (gene set-2-red sphere)) at right side corner. EEW63678.1 (Mga-HTH superfamily-blue sphere) solely interacted with EEW63479.1 (Accession_no. WP_002287056.1; Bacterial Ig-like domain-1 (gene set-2-grey sphere)) at 6 O' clock position Same set of proteins (78 HPs) were subjected to search against DEG database. It showed hit for 77 HPs and considered as the essential HPs (EHPs). These 77 EHPs were subjected to protein BLAST (BLASTp) search against Homo sapiens (taxid: 9606) proteome of the non-redundant database with significant parameters. Out of which, 65 proteins were exclusively present on pathogen and predicted as the potential drug targets. While excluding the significant hits, out of 5 adhesin, 2 showed similar match with human proteome (excluded from the list subsequently) and the residual were considered as novel targets.

peptide deformylase, ribokinase, sialidase-A, telomere resolvase rest, topoisomerase-IV-subunit-A, toxin-A, UDP-N-acetylmuramoyl-Alanine-D-glutamate-ligase). The rest of the reputed drug target candidates can be considered as novel targets, which need further experimental validation.

In a total of 23 potential drug targets, 5 are adhesins. Adhesins are one of the significant virulent factor involved in causing infection. Out of 5 adhesins, 3 (WP_002287056.1, WP_002286311.1 and WP_002340339.1) were exclusively present on the surface of the pathogen that can be targeted for therapeutic applications as penicillin binding protein4, uracil phosphoribosyl transferase and botulinum neurotoxin Type-F. Plus, 2 adhesins (WP_002287056.1 and WP_002340339.1) sequence and structures were similar to Big-1 domain adhesins. Via, I-TASSER and PROCHECK server 3D models for the three virulent adhesins were predicted and validated. The structure prediction and validation details are given in the Table 2. The validated structures were submitted to the protein model database with the following PMDB_ID PM0082281, PM0082280 and PM0082282.

DISCUSSION

The function of HPs encoded by protein coding genes are unknown hence this study attempts to open up a path to functional characterization of these proteins. In a total of 204 HPs, 190 HPs functions were predicted by different bioinformatics tools, among them 78 were predicted with high confidence and 27 enzymatic activities including adhesins.

All HPs including adhesins predicted to contain significant thermostability, hydropathicity, MW and Isoelectric point and instability index confirms the pathogenicity of the organism. As reported for *Haemophilus influenza*, high molecular weight adhesins attributed to human epithelial binding.[37] Similarly, adhesion (accession_no.WP_002340339.1) has highest MW, possibly supports human

Table 1: Functional assignment and classification and βαβ motifs (secondary structure) prediction.

HPs	DOMAIN/MOTIF				Classification	Secondary structure prediction	
Accession_no.	PSI-BLAST	Pfam	Interpro	CDD	Function	βαβ motifs	Range
WP_002286213.1	AAA_27_super_family	AAA-domain	AAA-domain	AAA_27_super_family	ATPase	–	
WP_000675717.1	Abi_C_superfamily	Abortive_infection_C-terminus	Abi_C	Abi_C_superfamily	APRP	–	
WP_002322465.1	AbiH_superfamily	Bacteriophage_abortive_infection_AbiH	AbiH	Bacteriophage_abortive_infection	APRP	–	
WP_002350625.1	APH-super_family and Capsule_synthesis_protein	Unknown	Nucleotide-diphospho-sugar transferases	APH super family	SRP	–	eeeeecchhh hhccceeeee 606 to 625 (19)
WP_002296543.1	ArnT_super_family	Unknown	Unknown	ArnT_super_family	translocase		
WP_002298929.1	B3_4_super_family	B3/4_domain	B3/B4_tRNA-binding_domain	B3_4_super_family	ligase	–	
WP_002287056.1	Big_3 super_family	Bacterial(Ig-like)domain3	Immunoglobulin-like	Big_3	Adhesin	–	
WP_002341586.1	Citrate_transporter	Citrate_transporter	Citrate_transporter-like_domain	Unknown	TP	–	
WP_002321275.1	Class_C_sortase	Gram-positive_pilin_subunitD1	Grampositive_pilin_subuniD1	Unknown	Adhesin	–	
WP_002340339.1	COG3942_superfamily	Bacterial-Ig-domain	Bacterial-Ig-domain	COG3942_superfamily	Adhesin	–	

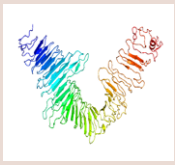
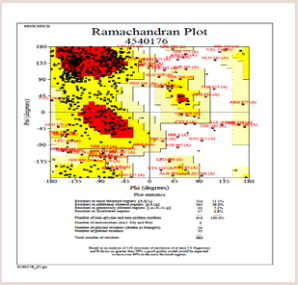
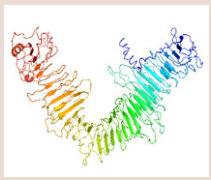
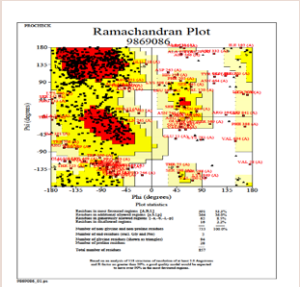
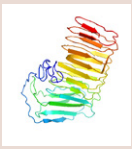
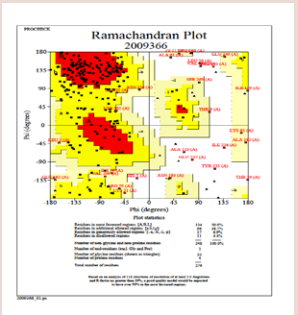
WP_002325891.1	COG3942_super family	NlpC/P60_ family&Mannosyl-glycoprotein-endo-beta-N-acetylglucosaminidase	Endopeptidase, NLPC/P60 domain	COG3942+NLPC_ P60&Spr-superfamily	hydrolase	_	
WP_002295906.1	CwlO1_superfamily	Unknown	Unknown	CwlO1	hydrolase	_	
WP_002292340.1	Cyclic-di-AMP_receptor	Cyclic-di-AMP_receptor	Cyclic-di-AMPreceptor	Cyclic-di-AMPreceptor	extracellular signalling protein	_	
WP_098381460.1	DDE_Tnp_ISL3_ superfamily	Transposase	Transposase	DDE_Tnp_ISL3 superfamily	Transposase	eeeeeeccccchhhhh hhhccccceee	49 to 72 (23)
WP_000185761.1	Protein_mom	Unknown	DNA_modification_protein	Unknown	transferase	_	
WP_002289495.1	DUF2807&SMC_N	DUF4097Putative-adhesin	Putative-adhesin	DUF2807-superfamily&SMC_N-superfamily	Adhesin	eeeeeeeeccccchhh hhhhhhccceeeee	330 to 360 (30)
WP_002325767.1	FAD_ binding_2superfamily	Unknown	Unknown	FAD_ binding_2superfamily	BP	_	
WP_106018980.1	FliI_super_family	MobL_relaxases	MobL_relaxases	FliI_super_family	relaxase	_	
WP_002289292.1	G6PD_bac	G6PD_bact	Glucose-6-phosphate-1-dehydrogenase,bacterial	Glucose-6-phosphate-1-dehydrogenase	oxidation-reduction	_	
WP_002326717.1	GDSL_family_lipase	Unknown	Acyl-CoA_N-acyltransferase	Unknown	transferase	_	
WP_106018971.1	Glucosaminidase_ super_family	Glucosaminidase	Mannosyl-glycoprotein_endo-beta-N-acetylglucosamidase-like-domain	Glucosaminidase_super_ family	hydrolase	eeeeeeccccchhhhh hhhhccccccccceeee	215 to 245 (30)
WP_002288314.1	Glycine-rich_SFCGS_ superfamily	Glycine-rich_SFCGS	HP	Glycine-rich_SFCGS_ superfamily	unknown	eeeeeeccccccchhhh hhhhccccceee	2 to 30 (28)
WP_002350626.1	Glyco_tranf_GTA_type	2-C-methyl-D-erythritol-4-phosphate_ cytidylyltransferase	Cytidylyltransferase-IspD/ TarI&BcbE,	GT2_BcE_like	transferase	_	
WP_002335549.1	Glyphos_transf	Glyphos_transf	CDP-glycerol-glycerophosphotransferase	Glyphos_transf_ superfamily	transferase	_	
WP_002286831.1	GyrI-like_super_family	GyrI-like_small_molecule_ binding_domain	Regulator_ factor	GyrI	BP	_	
WP_002352495.1	Halogen_Hydrol_ superfamily	Halogen_Hydrol	5-bromo-4-chloroindolyl_ phosphate	Halogen_Hydrol	hydrolysis	_	
WP_002296832.1	HelD_superfamily	AAA-domain	DNA-helicase	DNA helicaseIV	helicase	_	
WP_002286553.1	HNHc_6-superfamily&Lon-superfamily	Putative_HNHc-nuclease	Putative_HNHc-nuclease	HNHc_6-superfamily&Lon-superfamily	nuclease	_	
WP_002288760.1	HP	baeRF_family6	baeRF_family6	unknown	SRP	eeeeeccccchhhh hhhhccccccee	228 to 252 (24)
WP_002339568.1	alpha/beta-hydrolase/ DEAD/DEAH-box-helicase	DUF5348	DUF5348	Unknown	unknown	_	
WP_002286311.1	Putative-adhesin	Putative-adhesin	Putative-adhesin	Unknown	Adhesin	_	
WP_106018948.1	Beta-hexosaminidase	DUF5596	DUF5596	Unknown	unknown	_	
WP_002314399.1	HP	Family of unknown function (DUF5406)	Unknown	Unknown	unknown	_	

WP_002321269.1	DEAD/DEAH-box_ helicase	DUF5406	DUF5406	Unknown	helicase	_	
WP_002297506.1	pXO2-11	DUF5592	DUF5592	Unknown	unknown	_	
WP_002287057.1	HTH-superfamily	MgaHTH-superfamily	Mga-helix-turn-helix domain	HTH-superfamily	BP	_	
WP_002311258.1	HTH super family	Mga-HTH super family	Mga-HTH super family	Mga-HTH super family	BP	_	
WP_002287140.1	HTH-superfamily	MgaHTH-superfamily	MgaHTH-superfamily	HTH-superfamily	BP	_	
WP_002305295.1	HTH_40superfamily	HTH-superfamily	HTH-superfamily	HTH-superfamily	BP	_	
WP_106018943.1	HTH_Tnp_1&rve_3_ superfamily	HTH_Tnp_1&rve_2	Homeobox-like-domain_ superfamily	HTH_ Tnp_1&rve_3superfamily	BP	_	
WP_002286495.1	IFT57superfamily	Unknown	Unknown	Intra-flagellar_ transportprotein57	TP	_	
WP_002286016.1	LanC_like-superfamily	Glycosyl_Hydrolase-Family88	Six-hairpin-glycosidase-like-superfamily	LanC_like-superfamily	hydrolase	_	
WP_002345015.1	lectin_L-type&MucBP	Bacterial_lectin&MucBP	MucBP_domain	lectin_L-type&MucBP	Adhesin	eeeeeeeccccccccchhhh hhhccccceeeeee	630 to 652 (22)
WP_002287073.1	LURP-one-related-superfamily	Unknown	Homologous-superfamilyIPR025659Tubby-like,C-terminal	LOR-superfamily	IP	_	
WP_002303352.1	lysozyme_like super family	Lysozyme-like	Lysozyme-like	lysozyme_like	hydrolase	_	
WP_002296632.1	M-protein_transacting_ positive_regulator	Mga_helix-turn-helix_ domain	Mga_helix-turn-helix_domain	HTH-superfamily	BP	_	
WP_002335550.1	MATE_like_superfamily	Polysacc_synt&Polysacc_ synt_C	Polysaccharide_biosynthesis_ protein	MATE_like_superfamily	TP	_	
WP_002317290.1	MT-A70	MT-A70	methyltransferase	MT-A70_superfamily	transferase	_	
WP_002290045.1	NTP-PPase_BsYpjD	MazG-nucleotide_ pyrophosphohydrolase	MazG-nucleotide_ pyrophosphohydrolase	NTP-PPase_BsYpjD	hydrolase	_	
WP_002300053.1	Pentapeptide_4_ superfamily	Pentapeptide_repeats	Pentapeptide_repeat	Pentapeptide_4_ superfamily	unknown	_	
WP_002347500.1	Peptidase_C39like-family	Peptidase_C39like-family	DomainIPR039564-Peptidase-C39-like	Peptidase-familyC39	IP	_	
WP_002287397.1	PG_binding_4,PG_ binding_4&YkuD	PG_binding_4,PG_ binding_4&YkuD	L,D-transpeptidase	PG_binding_4,PG_ binding_4&YkuD	hydrolase	eeeccccccccccccchhhhh hhhhhcccceee	434 to 466 (32)
WP_106018942.1	PMT_2_superfamily	Unknown	Unknown	PMT_2_superfamily	transferase	_	
WP_002306002.1	PNPOx/FlaRed_like_ superfamily	Unknown	FMN-binding_splitbarrel	PNPOx/FlaRed_like_ superfamily	oxidoreductase	_	
WP_002325481.1	polymerase	Wzy_C-O-Antigen_ligase	O-antigen_ligase-related	unknown	ligase	_	
WP_002295260.1	PP/heavy_metal-binding_protein	Unknown	UPF0145-superfamily	Unknown	MiP	_	
WP_002296595.1	Prophage_tail-superfamily	Prophage-endopeptidase	Prophage_tail-endopeptidase	Prophage_tail-superfamily	hydrolase	eeccccccch hhhccceee	612 to 630 (18)
WP_002296335.1	RNase_PH-superfamily	Unknown	Unknown	RNase_PH-superfamily	transferase	_	
WP_077828678.1	Rve_superfamily-transposase	Integrase-coredomain	Transposase	Rve_superfamily-Integrase	Transposase	_	
WP_002322761.1	SdpI-family-protein	SdpI/YhfL-family	SdpI/YhfL-family	Unknown	IP	_	

WP_002341521.1	site-specific_integrase	Phage_int_SAM_3-family	Integrase/recombinase,N-terminal	Unknown	integrase	_	
WP_002296429.1	SMC_N_superfamily	Unknown	Unknown	SMC_N_superfamily	BP	_	
WP_002354430.1	SMC_N_superfamily	Unknown	Unknown	SMC_N_superfamily	BP	_	
WP_002287480.1	SWIM-superfamily	Unknown	SWIM-type_protein	SWIM-superfamily	MiP	_	
WP_002341874.1	Tim44-superfamily	Tim44-superfamily	Tim44-superfamily	Tim44-superfamily	translocase	_	
WP_002287659.1	TnpB_IS66-superfamily	IS66_Orf2-like-protein	Transposase-IS66	TnpB_IS66-superfamily	Transposase	_	
WP_002340450.1	TPP_enzyme_PYR_ superfamily	Unknown	Unknown	TPP_enzyme_PYR_ superfamily	BP	_	
WP_002341445.1	TRP	Unknown	Winged_helix-like-DNA-binding_domain	unknown	BP	_	
WP_106018950.1	Transposase_mut-superfamily	Transposase_mut-superfamily	Transposase_mut-superfamily	Transposase_mut-superfamily	Transposase	eecccccccccchhhh hhhccccccceeeeee	141 to 171 (30)
WP_106018970.1	tRNA_SAD-superfamily	Unknown	Uncharacterized-protein	tRNA_SAD-superfamily	ligase	_	
WP_002340448.1	WavE-superfamily	WavE-superfamily	WavE-superfamily	WavE-superfamily	BP	eeeeeeeeccccchhhh hhhhhccceeeee	109 to 136 (27)
WP_002287822.1	YbbR-superfamily	YbbR-like-protein	YbbR-like	YbbR-superfamily	unknown	_	
WP_002340505.1	YfbU-superfamily	YfbU-superfamily	YfbU	YfbU	unknown	_	
WP_002287574.1	YrhK_superfamily	YrhK-like_protein	DomainIPR025424-YrhK-domain	YrhK-like_protein	unknown	_	
WP_002288852.1	YycH-superfamily	YycH-protein	YycH	YycH-superfamily	MiP	_	
WP_002288853.1	YycI-like-superfamily	YycH_protein	YycH_protein	YycH-like	MiP	eeeccchhhh hhhccccee	36 to 56 (20)
WP_002295486.1	CAP_assoc_N superfamily	CAP-associated_N-terminal	CAP_superfamily	CAP-associated_N-terminal	hydrolase	_	
WP_101706209.1	Neuromodulin_N_ superfamily	Unknown	Unknown	Neuromodulin_N_ superfamily	BP	_	

*BP-Binding protein; Mip-Metal ion binding protein; TP-Transport Protein; IP-Immunity Protein; SRP-Stress Response Protein; TRP-Transcription Regulator Protein; APRP- Abortive Phage Resistance Protein; PP-Predicted protein

Table 2: I-TASSER Models for three virulent adhesins with structure evaluation.

s.no	Models (PMDB_ID)	5 analogs (PDB_ID)	Parameters	Ligand name	Ligand binding Residues	PROCHECK Evaluation
1.	 PM0082281	5n8pA 5gr8A 5hyxB 4ecnA 5gijB	C-score= -0.47 TM-score = 0.65±0.13 RMSD = 9.1±4.6Å	NAG MG NO CA CA	ASP216, THR217, ALA234, LYS236 ASP176, ILE177, ASP216 VAL215, GLN260 VAL316, GLU318 LYS256, GLU318	
2.	 PM0082280	5n8pA 5gr8A 5hyxB 2a0zA 3cigA	C-score = -1.67 TM-score = 0.51±0.15 RMSD = 12.7±4.3Å	I3C 2AN CA RX8 MG	VAL370, VAL372, GLN396, ILE397, HIS398, GLY420, MET421, SER422, ASP440, ASP442 GLN485, GLY486 GLU50, ASP101, ASP287 GLY359, SER368, ARG369, VAL387 THR105, ASP144	
3.	 PM0082282	3jx8A 3lycA 3ljyA 3petA 4y9vA1	C-score= -1.45 TM-score = 0.54±0.15 RMSD = 9.3±4.6Å	MG MAN XE GOL 3LJYB00	GLU237, LEU256, ASP276 SER194, ASN195 LEU20, SER137, GLY156, TRP157 ARG178, HIS197, GLU199 SER67, ILU69, GLU75, TYR118, ILU119, VAL120, ASN142	

epithelial binding. Adhesin's thermostability plays a major role in the compactness of the β-barrel like IgG and it also increases the resistant against proteases.[5]

The virulence factors are the key to design new anti-virulent drugs to prevent the host from the infection.[38] 79% HPs were predicted to be virulent and majority of them reside in the cytoplasm and cell membrane including seven Ig-like domain containing adhesins. According to Prava (2018), cytoplasmic proteins are probable drug target and membrane proteins are the potential vaccine target.[15] The ideal fixed length for the reported βαβ motifs pattern was maximum up to 32. Approximately, 8% HPs including two adhesin (WP_002289495.1 and WP_002345015.1) proteins have this pattern possibly contain active site.

According to Szklarczyk (2015), specific and productive functional relationship between two proteins, likely contributing to a common biological purpose.[39,40] Likewise, the PPI network of adhesin's gene sets possibly have the common biological role [Figure 3]. Followed by PPI, essential human non-homologous and novel drug targets were predicted.

Ultimately, the three therapeutic targets adhesin 3D structures were generated using I-TASSER. In general, they are beta-sheet comprising domains belongs to intimin/invasion family of bacteria, involved in initial contact with mammalian host cell through disulfide bond. And are commonly known as Big-1 domain named after found in enteropathogenic *Escherichia coli* intimin and in *Yersinia pseudotuberculosis* invasion.[41] Similarly, HP-adhesin (WP_002287056.1) is found to contain cysteine residues and subsequently belongs to intimin family. The other type of adhesins capable of forming an Ig-like topology and are belong to the group of cell surface non-fimbrial adhesins with the adhesion function even in the absence of disulfide bond.[42] Correspondingly, the Ig-like domains of these adhesins (WP_002286311.1 and WP_002340339.1) possibly belong to the group of non-fimbrial or afimbrial adhesins, due to lacks disulfide bond. And all three adhesins structures were submitted to PMDB.

CONCLUSION

The comprehensive function of hypothetical proteins is important owing to immune-informatics and therapeutic science applications. The present work unveils the function of 204 HPs from *E. faecium* strain 13-022. All-inclusive analysis supports the understanding of HPs characteristics. The structural analysis of predicted adhesins showed high sequence and structural similarity with Big-1 domain and annotated to predict the active drug binding site with pathogenicity. Our findings may open up the better investigation of Big-1 domain of afimbrial adhesins as well as to find prior drugs for targeting *E. faecium* and other biotechnological applications such as targeting biofilm formation, quorum sensing and quorum quenching.

ACKNOWLEDGEMENT

We thank the Centre for Bioinformatics, Pondicherry University for the computational facilities to carry out this work. We also thank Mr. Pranavathiyani G (BICPU) and Mr. Karamveer (BICPU) for their valuable suggestions and support throughout this work.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS

Big-1: Bacterial immunoglobulin-like (Ig) domain 1; **CDD:** Conserved Domain Database; **DEG:** Database of Essential Genes; **DSSP:** Secondary Structure assignment for all Protein entries; **EHPs:** Essential hypothetical proteins; **GO:** Gene Ontology; **GRAVY:** Grand average hydropathy; **HPs:** Hypothetical Proteins; **PPI:** Protein-Protein interactions; **PSI-BLAST:** Position-Specific Iterative Basic Local Alignment Search Tool; **PSSM:** Position Specific Scoring Matrix; **SOPM:** Self-Optimized Method for protein secondary structure prediction; **SOPMA:** Self-Optimized Prediction Method with Alignment; **SVM:** Support Vector Machine.

REFERENCES

1. Dubin K, Pamer EG. Enterococci and Their Interactions with the Intestinal Microbiome. Microbiol Spectr. 2014;5(6):1-24.
2. Gao W, Howden BP, Stinear TP. Evolution of virulence in *Enterococcus faecium*, a hospital-adapted opportunistic pathogen. Curr Opin Microbiol. 2018;41:76-82.
3. Hancock LE, Murray BE, Sillanpää J. Enterococci: From Commensals to Leading Causes of Drug Resistant Infection. Enterococcal Cell Wall Components Struct. 2014.
4. Salmanov AG, Vdovychenko SY, Litus OI, Litus VI, Bisyuk YA, Bondarenko TM, *et al.* Prevalence of health care–associated infections and antimicrobial resistance of the responsible pathogens in Ukraine: Results of a multicenter study (2014-2016). Am J Infect Control. 2019;47(6):e15-20.
5. Vengadesan K, Narayana SVL. Structural biology of Gram-positive bacterial adhesins. Protein Science. 2011;20(5):759-72.
6. Haiko J, Westerlund-Wikström B. The role of the bacterial flagellum in adhesion and virulence. Biology. 2013;2(4):1242-67.
7. Wilson JW, Schurr MJ, LeBlanc CL, Ramamurthy R, Buchanan KL, Nickerson CA. Mechanisms of bacterial pathogenicity. Postgrad Med J. 2002;78(918):216-24.
8. Berne C, Ducret A, Hardy GG, Brun YV. Adhesins Involved in Attachment to Abiotic Surfaces by Gram-Negative Bacteria. Microbiol Spectr. 2015;3(4):10.1128/microbiolspec.MB-0018–2015.
9. Klemm P, Schembri M. Bacterial adhesins: Function and structure. International Journal of Medical Microbiology. 2000;290(1):27-35.
10. Willems RJ, Schaik WV. Transition of *Enterococcus faecium* from commensal organism to nosocomial pathogen. Future Microbiol. 2009;4(9):1125-35.
11. Kan A, Del VI, Rudge T, Federici F, Haseloff J. Intercellular adhesion promotes clonal mixing in growing bacterial populations. J R Soc Interface. 2018;15(146):20180406.
12. Ochoa SA, Escalona G, Cruz-Córdova A, Dávila LB, Saldaña Z, Cázares-Domínguez V, *et al.* Molecular analysis and distribution of multidrug-resistant *Enterococcus faecium* isolates belonging to clonal complex 17 in a tertiary care center in Mexico City. BMC Microbiol. 2013;13(1):291.
13. Lebreton F, Schaik WV, McGuire AM, Godfrey P, Griggs A, Mazumdar V, *et al.* Emergence of epidemic multidrug-resistant *Enterococcus faecium* from animal and commensal strains. M Bio. 2013;4(4):e00534-13.
14. Sivashankari S, Shanmughavel P. Functional annotation of hypothetical proteins: A review. Bioinformation. 2006;1(8):335-8.
15. Prava JGP, Pan A. Functional assignment for essential hypothetical proteins of *Staphylococcus aureus* N315. Int J Biol Macromol. 2018;108:765-74.
16. DaCosta WLO, Araujo CL, De A, Dias LM, DePereira LCS, Alves JTC, *et al.* Functional annotation of hypothetical proteins from the *Exiguobacterium antarcticum* strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. PLoS One. 2018;13(6):e0198965.
17. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, *et al.* CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017;45(D1):D200-3.
18. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, *et al.* The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47(D1):D427-32.
19. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, *et al.* InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236-40.
20. Guruprasad K, Reddy B, Pandit M. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting *in vivo* stability of a protein from its primary sequence. Protein Engineering. 1991;4(2):155-61.
21. Ikai A. Thermostability and Aliphatic Index of Globular Proteins. Journal of Biochemistry. 1981;88(6):1895-8.
22. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105-32.
23. Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, *et al.* CELLO2GO: A web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. PLoS One. 2014;9(6):e99368.
24. Shen HB, Chou KC. Gpos-mPLoc: A top-down approach to improve the quality of predicting subcellular localization of Gram-positive bacterial proteins. Protein Pept Lett. 2009;16(12):1478-84.
25. Saha S, Raghava GPS. VICMpred: An SVM-based method for the prediction of functional proteins of Gram-negative bacteria using amino acid patterns and composition. Genomics Proteomics Bioinformatics. 2006;4(1):42-7.
26. Garg A, Gupta D. Virulent Pred: A SVM based prediction method for virulent proteins in bacterial pathogens. BMC Bioinformatics. 2008;9(1):62.
27. Sun L, Hu X, Li S, Jiang Z, Li K. Prediction of complex super-secondary structure βαβ motifs based on combined features. Saudi J Biol Sci. 2016;23(1):66-71.
28. Kouza M, Faraggi E, Kolinski A, Kloczkowski A. The GOR Method of Protein Secondary Structure Prediction and Its Application as a Protein Aggregation Prediction Tool. Methods Mol Biol. 2017;1484:7-24.
29. Frishman D, Argos P. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. Protein Eng. 1996;9(2):133-42.
30. Geourjon C, Deleage G. SOPM: A self-optimized method for protein secondary structure prediction. Protein Eng. 1994;7(2):157-64.
31. Geourjon C, Deleage G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci. 1995;11(6):681-4.
32. Snel B, Lehmann G, Bork P, Huynen MA. STRING: A web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res. 2000;28(18):3442-4.
33. Zhang R, Ou HY, Zhang CT. DEG: A database of essential genes. Nucleic Acids Res. 2004;32(Database issue):D271-2.
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.
35. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, *et al.* The ChEMBL database in 2017. Nucleic Acids Res. 2017;45(D1):D945-54.
36. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: Protein structure and function prediction. Nature methods. United States. 2015;12(1):7-8.
37. StGeme JWR, Falkow S, Barenkamp TJS. High-Molecular-Weight Proteins of Nontypable Haemophilus influenzae Mediate Attachment to Human Epithelial Cells. Proceedings of the National Academy of Sciences of the United States of America. 1993;90(7):2875-9.
38. Allen RC, Popat R, Diggle SP, Brown SP. Targeting virulence: Can we make evolution-proof drugs?. Nature reviews. Microbiology. England. 2014;12(1):300-8.
39. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, *et al.* STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(D1):D447-52.
40. Vega LA, Malke H, McIver KS. Virulence-Related Transcriptional Regulators of *Streptococcus pyogenes*. Oklahoma City (OK). 2016.
41. Batchelor M, Prasannan S, Daniell S, Reece S, Connerton I, Bloomberg G, *et al.* Structural basis for recognition of the translocated intimin receptor (Tir) by intimin from enteropathogenic *Escherichia coli*. EMBO J. 2000;19(11):2452-64.
42. Bodelón G, Palomino C, Fernández LÁ. Immunoglobulin domains in *Escherichia coli* and other enterobacteria: From pathogenesis to applications in antibody technologies. FEMS Microbiol Rev. 2013;37(2)204-50.

Article History: Submission Date : 23-11-2020; Revised Date : 09-12-2020; Acceptance Date : 09-01-2021.
Cite this article: Pownraj B, Ramu M, Haque E. Important Insight of Hypothetical Proteins from *E. faecium* Strain 13-022, the Drug Targeted Virulent Big-1 Adhesins: An *in silco* Approach. Int. J. Pharm. Investigation. 2021;11(1):32-40.

Table S1: Functional assignment of HPs from *E. faecium* strain_13-022.

Accession no.	psi-blast	pfam	interpro	CDD	Functional assignment
WP_002296358.1	2-phosphosulfolactate-phosphatase	Unknown	Unknown	Unknown	hydrolase
WP_002342384.1	30S-ribosomal_protein	Unknown	Unknown	Unknown	BP
WP_002294410.1	3-dehydroquinate-synthase	unknown	unknown	unknown	lyase
WP_002295869.1	ABC-transporter	Unknown	Unknown	Unknown	TP
WP_106018956.1	ABC_transporter	Unknown	Unknown	Unknown	TP
WP_002368121.1	acetyl-CoA_synthetase-like-protein/M20/M25/ M40-family_metallo-hydrolase	Unknown	Unknown	Unknown	hydrolase
WP_002350628.1	acyltransferase	Unknown	Unknown	Unknown	transferase
WP_002298250.1	Alkane_1-monooxygenase	Unknown	Unknown	Unknown	oxidoreductases
WP_002289272.1	aminopeptidase P family protein	Unknown	Unknown	Unknown	hydrolase
WP_002293828.1	ATP-binding_cassette	Unknown	Unknown	Unknown	BP
WP_002289233.1	ATP-binding_protein	unknown	unknown	unknown	BP
WP_002286840.1	ATP-dependent_Clp-protease	Unknown	Unknown	Unknown	BP
WP_002287453.1	ATP-dependent_helicase	Unknown	Unknown	Unknown	helicase
WP_074400044.1	beta-carotene_15,15'-monooxygenase	Unknown	Unknown	Unknown	oxidoreductase
WP_002321530.1	Class_F_sortase	Unknown	Unknown	Unknown	Adhesin
WP_002293655.1	ComGG	Unknown	Unknown	Unknown	BP
WP_002288325.1	CsbD-family	Unknown	Unknown	Unknown	SRP
WP_002302842.1	cyclase	Unknown	Unknown	Unknown	cyclase
WP_002289403.1	Cytoplasmic_protein	Unknown	Unknown	Unknown	unknown
WP_049143547.1	Cytosine_deaminase	unknown	unknown	unknown	hydrolase
WP_002347494.1	Dna-mismatch-repairprotein	Unknown	Unknown	Unknown	BP
WP_010730972.1	DNA-primase	Unknown	Unknown	Unknown	transferase
WP_002303824.1	DNA-repair-protein-MmcB-related-protein	Unknown	Unknown	Unknown	nuclease
WP_002321715.1	DNA-bindingprotein	Unknown	Unknown	Unknown	BP
WP_002295807.1	DUF1269	Unknown	Unknown	Unknown	unknown
WP_002296227.1	DUF1642	Unknown	Unknown	Unknown	unknown
WP_002296499.1	DUF1998	Unknown	Unknown	Unknown	unknown
WP_002302142.1	DUF2975	Unknown	Unknown	Unknown	unknown
WP_049143544.1	DUF3221	Unknown	Unknown	Unknown	unknown
WP_002317282.1	DUF3848	Unknown	Unknown	Unknown	unknown
WP_002287080.1	DUF3895	unknown	unknown	Unknown	unknown
WP_002303842.1	DUF4767	Unknown	Unknown	Unknown	unknown
WP_002326255.1	FtsX-like-permease	Unknown	Unknown	Unknown	permease
WP_002289445.1	glucosyltransferase	unknown	unknown	unknown	transferase
WP_002286559.1	HK97-family	Unknown	Unknown	Unknown	hydrolase
WP_000163792.1	Insertion_element_protein	Unknown	unknown	unknown	TP
WP_002287507.1	SirB	Unknown	Unknown	Unknown	transferase
WP_002299230.1	Isoprenyl_transferase	Unknown	Unknown	Unknown	transferase
WP_002295768.1	kinase	Unknown	Unknown	Unknown	transferase
WP_002331275.1	Leucine_Rich_repeat	Unknown	Unknown	Unknown	Oxidoreductase
WP_002326259.1	lipoprotein	Unknown	Unknown	Unknown	hydrolase
WP_010730973.1	L-ribulose-5-phosphate 4-epimerase	Unknown	Unknown	Unknown	Isomerase
WP_106018974.1	L-ribulose-5-phosphate 4-epimerase	Unknown	Unknown	Unknown	Isomerase
WP_002296631.1	M-protein_transacting_positive_regulator	Unknown	Unknown	Unknown	BP

WP_002344946.1	M-protein_transacting_positive_regulatorHTH/ DNA-binding_protein	unknown	unknown	unknown	BP
WP_002293998.1	Major_facilitator_superfamily domain-containing_ protein	Unknown	Unknown	Unknown	TP
WP_002350622.1	mannose-6-phosphate_isomerase	Unknown	Unknown	Unknown	Isomerase
WP_002321568.1	MBL_fold_kynureninase	Unknown	Unknown	Unknown	transferase
WP_010729510.1	MP	Unknown	Unknown	Unknown	BP
WP_002294011.1	MP	Unknown	Unknown	Unknown	BP
WP_002288350.1	MP	Unknown	Unknown	Unknown	BP
WP_002287053.1	MP	Unknown	Unknown	Unknown	BP
WP_002317291.1	MP	Unknown	Unknown	Unknown	BP
WP_002322202.1	MP	unknown	unknown	unknown	BP
WP_002296481.1	methyltransferase	Unknown	Unknown	Unknown	transferase
WP_002287147.1	YtxH_domain-containing_protein	Unknown	Unknown	Unknown	Miscellaneous
WP_106018969.1	myosin-11-like	Unknown	Unknown	Unknown	unknown
WP_002302096.1	Oligosaccharide_biosynthesis_protein-Alg14	Unknown	Unknown	Unknown	transferase
WP_002368120.1	PcfX-family	Unknown	Unknown	Unknown	miscellaneous
WP_002289550.1	penicillin-BP	Unknown	Unknown	Unknown	BP
WP_002323868.1	Peptidase_S8/XRE family	Unknown	Unknown	Unknown	hydrolase
WP_002286523.1	phage_gp6-like_head-tail_connector_protein	Unknown	Unknown	Unknown	hydrolase
WP_002303420.1	Phage-head-tail_adapter_protein	Unknown	Putative_ metalloenzymes	Unknown	hydrolase
WP_002286524.1	Phage_protein	Unknown	Unknown	Unknown	hydrolase
WP_002286512.1	Phage_tail_protein	Unknown	Unknown	Unknown	hydrolase
WP_002295913.1	Phage_tail_protein	Unknown	Unknown	Unknown	hydrolase
WP_002326819.1	PP	Unknown	Unknown	Unknown	PP
WP_002350624.1	PP	unknown	unknown	unknown	PP
WP_002299301.1	PP	unknown	unknown	unknown	PP
WP_002295864.1	PP	Unknown	Unknown	Unknown	PP
WP_002289257.1	PP/Glycoside hydrolase family 13 protein	Unknown	Unknown	Unknown	PP
WP_010730971.1	PP-DNA_mismatch_repair-ATPase	unknown	unknown	unknown	PP
WP_000248477.1	phosphatase	Unknown	Unknown	Unknown	hydrolase
WP_002321681.1	PutativeE3-ubiquitin-protein_ligase	Unknown	Unknown	Unknown	ligase
WP_002324247.1	prepilin-type-N-terminal-cleavage/methylation_ domain	Unknown	Unknown	Unknown	TP
WP_002289649.1	Putative-lipoprotein	Unknown	Unknown	Unknown	permease
WP_002295447.1	Putative-MP	Unknown	Unknown	Unknown	TP
WP_002321168.1	Putative-MP	Unknown	Unknown	Unknown	TP
WP_002340550.1	Putative-MP	unknown	unknown	unknown	TP
WP_002296549.1	Putative-MP	unknown	unknown	unknown	TP
WP_002303114.1	Putative-UV-damage_repair_protein	Unknown	Unknown	Unknown	transferase
WP_002347528.1	Putative-UV-damage_repair_protein	Unknown	Unknown	Unknown	transferase
WP_002350357.1	pXO2-10-like-protein	Unknown	Unknown	Unknown	Miscellaneous
WP_070828461.1	Restriction-endonuclease_subunitR	Unknown	Unknown	Unknown	BP
WP_002288643.1	ribosomal-protein-alanine-N-acetyltransferase	Unknown	Unknown	Unknown	oxidoreductase
WP_002301068.1	RnfC_N-superfamily	Unknown	Unknown	RnfC_N-superfamily	oxidoreductase
WP_002286049.1	rRNA_processing-protein	Unknown	Unknown	Unknown	transferase

WP_002330700.1	S-adenosyl-L-methionine-dependent-methyltransferase	Unknown	Unknown	Unknown	transferase
WP_002323892.1	SdpI-family-protein	Unknown	Unknown	Unknown	IP
WP_000455809.1	SEC-C_domain-containing_protein	Unknown	Unknown	Unknown	BP
WP_002287075.1	Sex-pheromone_cAM373	Unknown	Unknown	Unknown	unknown
WP_002346986.1	Signal-peptide	Unknown	Unknown	Unknown	translocase
WP_074400136.1	SIR2-family_protein/beta-N-acetylhexosaminidase	Unknown	Unknown	Unknown	unknown
WP_002285960.1	Sugar-transporter	Unknown	Unknown	Unknown	unknown
WP_106018965.1	Terminase_large-subunit	Unknown	Unknown	Unknown	unknown
WP_049143545.1	TetR/AcrR-family	Unknown	Unknown	Unknown	hydrolase
WP_002320994.1	Tetraacyldisaccharide-4'-kinase	Unknown	Unknown	Unknown	unknown
WP_002323140.1	Tetratricopeptide-repeat	Unknown	Unknown	Unknown	unknown
WP_010706480.1	TlpA-family-disulfide-reductase	Unknown	Unknown	Unknown	reductase
WP_002347493.1	TPR_MLP1_2-superfamily	unknown	unknown	TPR_MLP1_2-superfamily	BP
WP_002310954.1	TraE1-conjugal_transfer-protein	Unknown	Unknown	Unknown	transferase
WP_002304624.1	TRP	Unknown	Unknown	Unknown	BP
WP_074400096.1	TRP	Unknown	Unknown	Unknown	BP
WP_002350689.1	transposase	unknown	unknown	unknown	transposase
WP_002301170.1	Trichoplein-keratin-filament-BP	Unknown	Unknown	Unknown	BP
WP_002289192.1	translocase-TatC	Unknown	Unknown	Unknown	translocase
WP_002302159.1	permease/ATPase	Unknown	Unknown	Unknown	permease
WP_002320976.1	V-type-ATPase	Unknown	Unknown	Unknown	ATPase
WP_002340448.1	WavE-superfamily	WavE-superfamily	WavE-superfamily	WavE-superfamily	BP
WP_002296469.1	WxL domain surface protein	unknown	unknown	unknown	SRP&virulence
WP_002296825.1	XRE-family-TRP	Unknown	Unknown	Unknown	SRP&virulence
WP_002295457.1	XRE-family-TRP	Unknown	Unknown	Unknown	SRP&virulence
WP_002321678.1	YIP1-family	Unknown	Unknown	Unknown	TP
WP_002299629.1	Zinc-ribbon_domain	Unknown	Unknown	Unknown	unknown
WP_002286680.1	HP	Unknown	Unknown	Unknown	unknown
WP_002286695.1	HP	Unknown	Unknown	Unknown	unknown
WP_002286694.1	HP	Unknown	Unknown	Unknown	unknown
WP_002288892.1	HP	Unknown	Unknown	Unknown	unknown
WP_002311569.1	HP	Unknown	Unknown	Unknown	unknown
WP_002289611.1	HP	Unknown	Unknown	Unknown	unknown
WP_002292681.1	HP	Unknown	Unknown	Unknown	unknown
WP_002289266.1	HP	Unknown	Unknown	Unknown	unknown
WP_106018962.1	HP	Unknown	Unknown	Unknown	unknown
WP_059355966.1	HP	Unknown	Unknown	Unknown	unknown
WP_002295905.1	HP	unknown	unknown	unknown	unknown
WP_002299302.1	HP	unknown	unknown	unknown	unknown
WP_002288487.1	HP	unknown	unknown	unknown	unknown

*BP-Binding protein; Mip-Metal ion binding protein; TP-Transport Protein; IP-Immunity Protein; SRP-Stress Response Protein; TRP-Transcription Regulator Protein; APRP- Abortive Phage Resistance Protein; PP-Predicted protein; HP-Hypothetical Protein; DUF-Domain of Unknown Function; MP-Membrane Protein;

Table S2: Physico-chemical properties of HPs.

HPs	Protparam				
Accession_no.	MV	PI	GRAVY	INSTABILITY_INDEX	ALIPHATIC_INDEX
WP_002324247.1	11671.55	5	-0.13	47.14	99.4
WP_002286680.1	12267.11	9.05	-0.509	37.28	72.1
WP_002286695.1	11580.35	9.76	-0.778	25.43	73.3
WP_002296358.1	10502.45	9.99	0.92	14.07	109.3
WP_002286049.1	11273.04	4.8	-1.504	52.2	40.59
WP_002286694.1	11914.73	9.79	-0.835	49.18	79.21
WP_002296825.1	11877.74	9.3	-0.641	40.96	77.57
WP_002286523.1	11945.37	4.31	-0.35	60.4	91.84
WP_002296499.1	11724.22	4.04	-0.25	54.03	110.78
WP_002350622.1	11950.04	8.85	-0.176	34.85	86.02
WP_002331275.1	12516.08	4.27	-0.366	52.99	83.37
WP_002320994.1	12103.67	5.52	-0.774	46.64	67.4
WP_002288892.1	11983.29	10.61	0.328	40.21	109.71
WP_002323892.1	11829.89	9.4	0.604	44.35	125.58
WP_002295869.1	11983.33	10.09	0.346	21.31	104.95
WP_002322761.1	11881.18	10.4	0.712	47.64	140.19
WP_002323868.1	12460.14	6.17	-0.299	59.38	70.85
WP_002295768.1	12133.87	6.59	-0.678	48.35	74.62
WP_002287574.1	12601.87	10.04	-0.203	52.68	99.43
WP_002321530.1	12354.38	5.06	-0.326	13.84	97.45
WP_002321681.1	12234.05	9.85	1.317	24.2	161.21
WP_002340505.1	12941.45	4.99	-0.713	43.26	77.38
WP_002289192.1	12576.41	10.14	0.844	45.17	137.48
WP_002321715.1	12834.15	8.66	-0.371	47.52	91.2
WP_002287453.1	12687.24	4.19	-0.556	39.2	84.91
WP_002299230.1	12250.14	9.37	-0.712	55.09	93.06
WP_106018942.1	12743.07	9.34	-0.057	56.62	96.61
WP_098381460.1	12842.21	10.87	-0.272	29.23	84.04
WP_002292340.1	11870.61	4.77	0.077	39.56	102.75
WP_002296227.1	12511.43	4.72	-0.236	33.83	100.18
WP_002288325.1	12849.15	4.91	-1.646	66.3	55.91
WP_002297506.1	13113.23	7.89	0.03	27.81	88.73
WP_002310954.1	12537.9	9.7	-0.377	45.18	95.64
WP_002320976.1	13195.96	6.77	-1.355	57.4	75.45
WP_002290045.1	12900.51	4.97	-0.775	36.94	68.18
WP_002287147.1	12185.33	9.27	-0.113	16.91	89.46
WP_002302842.1	13300.03	4.61	-0.629	44	74.59
WP_002301068.1	12657.56	6.28	-0.189	54.08	98.38
WP_002295260.1	12904.79	5.38	-0.187	42.44	91.34
WP_002288643.1	12942.51	4.18	-0.384	45.93	93.04
WP_002306002.1	13067.95	5.4	-0.335	40.09	85.27
WP_077828678.1	12824.97	8.73	-0.063	25.7	99.2

WP_002341586.1	11926.46	9.45	1.057	31.14	128.5
WP_002326717.1	13168.26	9.19	-0.55	47.7	93.27
WP_002286559.1	13375.85	4.43	-0.781	33.41	90.62
WP_002314399.1	13083.46	4.11	-0.56	32.93	78.41
WP_002321269.1	13156.52	4.21	-0.677	34.15	71.5
WP_002368121.1	12949.99	6.22	-0.236	21	100
WP_002287507.1	13448.97	9.84	0.465	36.13	115.53
WP_002289550.1	13420.87	10	0.036	56.76	97.57
WP_002347528.1	13520.19	5.23	-0.69	49.54	85.57
WP_010706480.1	13604.52	5.64	-0.855	31.15	74.05
WP_002296335.1	13645.39	4.45	-0.38	51.65	105.95
WP_002303114.1	13599.99	4.71	-0.91	57.45	79.05
WP_002287659.1	13771.03	8.77	-0.343	30.85	101.62
WP_002321678.1	13583.52	8.97	-0.462	25.44	75.04
WP_002288314.1	12429.36	7.86	-0.223	22.76	71.95
WP_002296631.1	14051.16	4.74	-0.126	30.14	103.22
WP_002311569.1	13817.38	9.47	0.114	21.03	123.05
WP_002341521.1	14076.23	9.61	-0.874	20.86	85
WP_002293655.1	13794	5.41	0.097	53.3	93.36
WP_106018943.1	13964.04	9.07	-0.697	35.07	79.5
WP_002287075.1	19211.99	6.11	-0.417	20	79.7
WP_002339568.1	20132.07	5.68	-0.675	26.41	85.35
WP_070828461.1	20099.79	4.61	-0.323	42.08	92.28
WP_002287073.1	20168.14	8.76	-0.383	26.22	80.94
WP_002285960.1	20340.26	10	0.254	26.91	109.36
WP_002303352.1	19747.23	4.94	-0.53	32.28	81.21
WP_002295447.1	19271.88	4.82	0.664	42	117.64
WP_002289403.1	20320.52	5.77	-0.21	49	98
WP_002289611.1	20929.49	9.65	0.802	29.12	136.05
WP_002301170.1	19718.54	9.8	-0.236	33.83	66.48
WP_010729510.1	20323.51	9.3	-0.159	23.45	91.96
WP_002317290.1	20776.11	8.81	-0.167	46.46	88.89
WP_002292681.1	21071.09	9.5	-0.957	33.81	81.92
WP_002330700.1	21629.62	5.3	-0.299	48.42	101.64
WP_002295457.1	22261.29	4.92	-0.72	40.73	89.62
WP_002303842.1	21582.19	9.45	-0.34	49.09	88.66
WP_106018965.1	22565.43	4.93	-0.593	37.41	95.34
WP_002293998.1	22762.5	10	0.668	38.93	123.49
WP_002295807.1	21230.78	4.94	-0.42	30.33	81.54
WP_002350357.1	23520.2	9.42	-0.774	47.99	89.44
WP_002322465.1	23633.24	4.66	-0.523	45.21	80.25
WP_074400136.1	23389.85	4.6	-0.319	44.41	100.95
WP_002286512.1	22152.03	4.88	-0.286	20.25	76.73
WP_002286831.1	24129.01	8.95	-0.695	55.69	73.76
WP_002321568.1	23591.14	5.51	-0.067	52.55	99.76

WP_002289649.1	24551.94	9.13	0.856	26.69	132.27
WP_002295864.1	24074.62	5.14	-0.297	40.9	92.71
WP_002293828.1	23488.54	8.81	-0.605	19.82	76.33
WP_002296481.1	24389.66	4.83	-0.354	32.68	84.33
WP_002294011.1	23320.84	9.84	0.698	27.09	108.71
WP_002300053.1	23751.97	4.97	-0.107	32.33	86.38
WP_002289266.1	24247.64	8.71	0.745	32.22	108.19
WP_049143544.1	23373.27	5.1	-0.25	30.13	78.95
WP_049143545.1	24112.46	5.12	-0.383	39.42	99.76
WP_000248477.1	25050.43	5.82	-0.722	33.33	73.07
WP_106018970.1	24286.12	5.01	-0.766	28.4	77.9
WP_002350628.1	24543.51	9.47	0.851	18.98	140.19
WP_002298250.1	25517.4	5.46	-0.335	36.85	95.67
WP_002352495.1	25134.91	6	-0.437	43.07	98.89
WP_002299629.1	25583.38	4.77	0.059	44.11	104.17
WP_002303420.1	26048.95	6.11	-0.466	47.39	92.57
WP_002368120.1	25403.11	6.64	-0.471	33.56	85.02
WP_074400096.1	25975.69	9.47	-0.035	38.2	115.02
WP_002289292.1	26635.24	6.01	-0.868	48.61	73.8
WP_106018950.1	25437.06	9.02	-0.582	33.54	86.08
WP_002288350.1	25340.24	8.35	0.733	19.92	111.52
WP_106018980.1	27529.3	9.6	-1.318	60.82	64.89
WP_002286553.1	26647.8	9.12	-0.373	34.13	84.25
WP_002303824.1	26838.61	5.28	-0.571	61.04	80.87
WP_002302159.1	26616.63	5.13	-0.298	74.73	94.59
WP_002346986.1	24732.96	5.3	-0.214	29.12	72.84
WP_002347494.1	27332.83	7.04	-0.954	51.27	74.45
WP_002296429.1	26100.94	4.78	-0.89	29.37	86.57
WP_002326819.1	28122.47	4.63	-0.83	52.92	69.09
WP_002289272.1	26863.41	5.76	-0.699	39.15	82.52
WP_002298929.1	26411.57	4.67	-0.387	43.93	86.48
WP_002325767.1	25427.08	4.88	-0.33	46.29	74.43
WP_002350626.1	27259.08	4.85	-0.168	28.2	86.34
WP_002326255.1	28149.36	9.18	0.649	45.65	117.39
WP_002341874.1	27227.76	9.68	-0.63	45.01	60.08
WP_002289257.1	28753.09	9.5	-1.322	60.41	52.68
WP_074400044.1	28043.14	4.99	0.226	26.54	98.35
WP_002286840.1	29618.9	6.07	-0.498	32.8	83.84
WP_002302142.1	29856.17	5.73	-0.255	34.32	98.12
WP_000185761.1	30671.48	9.27	-0.518	36.33	78.6
WP_106018962.1	30307.74	6	-0.254	27.1	96.3
WP_000675717.1	29750.71	5.49	-0.425	34.32	88.8
WP_002304624.1	29063.87	5.23	-0.35	23.13	88.53
WP_106018971.1	28559.42	5.95	-0.1	45.49	74.64
WP_002326259.1	29222.77	5	-0.502	45.77	76.81

WP_010730972.1	31131.23	9.32	-0.948	34.93	65.76
WP_106018956.1	31494.9	5.63	-0.818	33.6	83.66
WP_002287053.1	30804.2	6.9	0.381	10.91	110.11
WP_106018969.1	31515.02	6.06	-0.796	29.45	84.74
WP_002340450.1	30755.01	4.89	-0.078	32.47	91.37
WP_106018948.1	32414.29	7.71	-0.402	36.67	84.05
WP_002321275.1	30819.8	4.58	-0.72	53.8	68.04
WP_002286311.1	30221.97	4.91	-0.388	11.89	95.36
WP_002342384.1	32072.65	7.69	-0.369	48.14	94.04
WP_002287480.1	33618.76	8.38	-0.49	41.7	84.04
WP_002317291.1	31094.64	4.65	0.786	25.44	118.54
WP_002288853.1	32958.77	4.75	-0.441	37.94	89.69
WP_002286524.1	34677.7	6.03	-0.454	32.94	86.36
WP_010730973.1	34370.76	9.52	-0.442	28.34	95.09
WP_002302096.1	35382.97	5.79	-0.616	45.66	75.64
WP_002295906.1	35966.8	4.62	-1.084	55.73	71.34
WP_002323140.1	36657.86	4.97	-0.473	59.82	90.03
WP_059355966.1	36346.01	8.59	0.182	38.85	110.87
WP_106018974.1	36499.18	9.65	-0.395	32.76	91.19
WP_002317282.1	36069.64	9.16	-0.704	45.66	71.18
WP_002295913.1	37036.48	6.98	-0.336	40.78	91.95
WP_002321168.1	38340.79	9.48	0.696	29.5	120.06
WP_000455809.1	38991.28	4.79	-0.435	57.6	88.11
WP_002305295.1	40152.82	5.37	-0.285	47.04	97.55
WP_002347500.1	39053.89	9.18	-0.469	28.25	66.77
WP_002294410.1	41047.52	5.72	-0.14	36.58	96.93
WP_002295905.1	41404.04	4.47	5.72	32.17	86.53
WP_002286016.1	43004.68	5.3	-0.549	35.7	79.75
WP_002288760.1	41783.08	4.98	-0.428	30.84	85.22
WP_010730971.1	43381.45	8.23	-0.609	33.44	71.02
WP_002295486.1	43317.91	4.7	-0.378	35.29	82.91
WP_002340448.1	43803.37	7.94	-0.303	34.15	84.26
WP_002335549.1	44729.61	9.14	-0.423	41.24	93.89
WP_002296632.1	46434.53	9.21	-0.545	43.35	80.03
WP_002287822.1	41389.78	4.79	-0.281	38.11	82.9
WP_002289233.1	44083.91	5.06	-0.35	34.14	77.28
WP_002299302.1	46378.07	9.52	0.493	30.9	117.53
WP_002350689.1	47844.06	9.61	-0.693	31.77	77.21
WP_002287080.1	51417.31	9.42	-0.7	43.17	79.36
WP_049143547.1	51936.19	9.04	-0.468	49.71	96.81
WP_002344946.1	51576.62	5.07	-0.312	46.05	90.32
WP_002288852.1	50036.18	5.14	-0.456	28.96	84.52
WP_002350624.1	50189.2	8.04	0.702	31.91	123.33
WP_002288487.1	53048.91	6.15	-0.516	40.56	87.71
WP_002296549.1	52491.76	9.53	0.712	31.9	128.01

WP_002325481.1	53315.37	9.49	0.621	22.85	118.12
WP_002287397.1	52056.24	5.04	-0.585	29.15	73.99
WP_002335550.1	54609.63	9.69	0.825	25.41	129.41
WP_002341445.1	57736.41	5.2	-0.229	47.65	104.3
WP_002287057.1	58772.17	7.19	-0.234	46.3	96.87
WP_002287140.1	58290.27	8.7	-0.345	43.83	93.73
WP_002311258.1	62111.65	6.18	-0.275	43.52	98.4
WP_002289495.1	58704.77	4.79	-0.683	32.68	86.32
WP_002347493.1	62709.11	7.1	-0.383	39.16	101.25
WP_002350625.1	65492.13	5.68	-0.297	34.1	93.4
WP_002322202.1	66336.21	9.36	0.343	38.26	108.83
WP_002289445.1	67721.27	9.4	0.2	27.89	107.58
WP_002299301.1	67610.94	9.34	0.536	31.4	112.77
WP_000163792.1	76746.29	9.28	-0.807	38.31	65.4
WP_002345015.1	74835.58	4.78	-0.403	31.78	87.39
WP_002296543.1	78530.75	7.09	0.27	30.98	104.79
WP_002354430.1	79929.47	4.68	-1.081	56.03	71.42
WP_002287056.1	76493.94	5.96	-0.597	46.2	77.99
WP_002286495.1	78321.65	5.11	-0.399	25.44	80.83
WP_002296832.1	82358.03	5.69	-0.458	43.25	95.72
WP_002325891.1	76931.22	4.81	-0.627	32.93	65.06
WP_002296595.1	84901.12	5.01	-0.351	18.09	83.42
WP_002296469.1	89435.15	5.3	-0.223	26.63	83.37
WP_101706209.1	92778.07	6.23	-0.437	29.93	88.16
WP_002340339.1	91739.41	4.66	-0.388	28.48	76.77
WP_002340550.1	99264.3	9.04	-0.46	28.76	79.67
WP_002286213.1	107021	5.74	-0.755	52.34	84.55

Table S3: Subcellular localization and virulence prediction of HPs.

HPs	Subcellular_localization		VF_prediction	
Accession_no.	[$]CELLO2GO	Gpos-mPLoc	VICMpred	VirulentPred
WP_002324247.1	Cytoplasmic	Extracell	MM	Virulent
WP_002286680.1	Cytoplasmic	CM	MM	Virulent
WP_002286695.1	Cytoplasmic	CM/Cytoplasm	CP	Virulent
WP_002296358.1	IM	CM	MM	Non-virulent
WP_002286049.1	Cytoplasmic	Extracell	CP	Virulent
WP_002286694.1	Cytoplasmic	Cytoplasm/Extracell	MM	Virulent
WP_002296825.1	Cytoplasmic	Extracell	CP	Non-virulent
WP_002286523.1	Cytoplasmic	Extracell	MM	Virulent
WP_002296499.1	Cytoplasmic	Cytoplasm	CP	Virulent
WP_002350622.1	Cytoplasmic	CM/Cytoplasm	CP	Virulent
WP_002331275.1	Cytoplasmic	CM	CP	Virulent
WP_002320994.1	Periplasmic/Cytoplasmic	CM/Extracell	CP	Virulent
WP_002288892.1	Cytoplasmic	CM	CP	Virulent
WP_002323892.1	IM	CM	CP	Virulent

WP_002295869.1	Periplasmic	CM	CP	Virulent
WP_002322761.1	IM	CM	MM	Virulent
WP_002323868.1	Periplasmic	CM	CP	Virulent
WP_002295768.1	Cytoplasmic	CM/Extracell	MM	Non-Virulent
WP_002287574.1	Cytoplasmic	CM	CP	Virulent
WP_002321530.1	Cytoplasmic	Extracell	CP	Non-Virulent
WP_002321681.1	IM	CM	MM	Virulent
WP_002340505.1	Cytoplasmic(PM1500)	CM	CP	Virulent
WP_002289192.1	IM	CM	MM	Virulent
WP_002321715.1	Cytoplasmic	CM/Extracell	MM	Virulent
WP_002287453.1	Cytoplasmic	Cytoplasm	CP	Virulent
WP_002299230.1	Cytoplasmic	Extracell	IS	Virulent
WP_106018942.1	Cytoplasmic	CM	MM	Virulent
WP_098381460.1	Periplasmic	CM/Extracell	MM	Virulent
WP_002292340.1	Cytoplasmic(yaaQ)	Cytoplasm	CP	Virulent
WP_002296227.1	Cytoplasmic	Extracell	IS	Virulent
WP_002288325.1	Cytoplasmic	Cytoplasm	CP	Virulent
WP_002297506.1	IM	CM	MM	Virulent
WP_002310954.1	Cytoplasmic	CM/Cytoplasm	MM	Virulent
WP_002320976.1	Cytoplasmic	CM/Cytoplasm	MM	Virulent
WP_002290045.1	Cytoplasmic(ypjD)	Cytoplasm	CP	Virulent
WP_002287147.1	Periplasmic	Extracell	CP	Virulent
WP_002302842.1	Cytoplasmic	CM/Cytoplasm	CP	Virulent
WP_002301068.1	Cytoplasmic/OM	CM	CP	Virulent
WP_002295260.1	Cytoplasmic	CM	VF	Virulent
WP_002288643.1	Cytoplasmic(ydfE)	Extracell	CP	Virulent
WP_002306002.1	Cytoplasmic(insF)	CM	CP	Virulent
WP_077828678.1	IM	CM/Cytoplasm	MM	Non-Virulent
WP_002341586.1	Cytoplasmic	CM	MM	Virulent
WP_002326717.1	Cytoplasmic	Extracell	CP	Virulent
WP_002286559.1	Cytoplasmic(EF_A0048)	Extracell	MM	Virulent
WP_002314399.1	Cytoplasmic(EF_A0048)	CM/CW	CP	Virulent
WP_002321269.1	Cytoplasmic	CM/Cytoplasm	CP	Virulent
WP_002368121.1	IM	Extracell	VF	Virulent
WP_002287507.1	Cytoplasmic	CM	MM	Virulent
WP_002289550.1	Cytoplasmic	CM	MM	Non-Virulent
WP_002347528.1	Cytoplasmic	Extracell	MM	Virulent
WP_010706480.1	Cytoplasmic	Extracell	CP	Virulent
WP_002296335.1	Cytoplasmic	Cytoplasm	IS	Non-Virulent
WP_002303114.1	Cytoplasmic(NGR_a03130)	Cytoplasm/Extracell	MM	Virulent
WP_002287659.1	Periplasmic	Cytoplasm	VF	Non-Virulent
WP_002321678.1	Cytoplasmic	CM	CP	Non-Virulent
WP_002288314.1	Cytoplasmic	Extracell	CP	Virulent
WP_002296631.1	Cytoplasmic	CM	MM	Virulent
WP_002311569.1	Cytoplasmic(ps201)	CM	MM	Virulent

WP_002341521.1	Cytoplasmic	CM/Extracell	CP	Virulent
WP_002293655.1	Cytoplasmic	CM	CP	Virulent
WP_106018943.1	Cytoplasmic	CM/Extracell	CP	Non-Virulent
WP_002287075.1	Cytoplasmic	Extracell	MM	Non-virulent
WP_002339568.1	Cytoplasmic	Extracell	CP	Virulent
WP_070828461.1	Cytoplasmic(yxjI)	CM	CP	Virulent
WP_002287073.1	IM	CM	MM	Virulent
WP_002285960.1	Cytoplasmic(yddH)	CM	CP	Virulent
WP_002303352.1	Cytoplasmic(yuaF)	Extracell	IS	Non-Virulent
WP_002295447.1	Cytoplasmic	CM	MM	Virulent
WP_002289403.1	IM	Extracell	CP	Virulent
WP_002289611.1	Extracellular	CM	CP	Virulent
WP_002301170.1	OM	Extracell	MM	Virulent
WP_010729510.1	Cytoplasmic(munIM)	Extracell	MM	Non-virulent
WP_002317290.1	Cytoplasmic	Cytoplasm	IS	Virulent
WP_002292681.1	Cytoplasmic	Cytoplasm/Extracell	MM	Virulent
WP_002330700.1	Cytoplasmic	CM	VF	Virulent
WP_002295457.1	Cytoplasmic	Extracell	CP	Virulent
WP_002303842.1	Cytoplasmic	CM/Cytoplasm	CP	Virulent
WP_106018965.1	IM	CM	CP	Non-virulent
WP_002293998.1	Cytoplasmic	CM	MM	Non-virulent
WP_002295807.1	Cytoplasmic(pXO2-10)	Extracell	CP	Virulent
WP_002350357.1	Cytoplasmic	CM/Extracell	MM	Virulent
WP_002322465.1	Cytoplasmic	CM	MM	Virulent
WP_074400136.1	Cytoplasmic	CM/Extracell	CP	Virulent
WP_002286512.1	Cytoplasmic	Extracell	CP	Non-virulent
WP_002286831.1	Cytoplasmic	Cytoplasm	CP	Virulent
WP_002321568.1	IM	Extracell	MM	Virulent
WP_002289649.1	Cytoplasmic	CM	CP	Non-virulent
WP_002295864.1	Extracellular	CM/Cytoplasm	CP	Virulent
WP_002293828.1	Cytoplasmic	Extracell	CP	Virulent
WP_002296481.1	IM(niaX)	CM	CP	Non-virulent
WP_002294011.1	Cytoplasmic(yisX)	CM	MM	Virulent
WP_002300053.1	IM	CM/Cytoplasm	CP	Virulent
WP_002289266.1	Extracellular	CM	CP	Virulent
WP_049143544.1	Cytoplasmic	Extracell	CP	Virulent
WP_049143545.1	OM	Extracell	CP	Virulent
WP_000248477.1	OM	CM	CP	Virulent
WP_106018970.1	IM	Extracell	CP	Virulent
WP_002350628.1	Cytoplasmic	CM	CP	Virulent
WP_002298250.1	Cytoplasmic(xpaC)	CM/Cytoplasm	CP	Virulent
WP_002352495.1	IM	CM/Cytoplasm.	CP	Virulent
WP_002299629.1	Cytoplasmic	CM	MM	Virulent
WP_002303420.1	Cytoplasmic	CM	MM	Virulent
WP_002368120.1	Cytoplasmic	CM/Extracell	CP	Virulent

WP_074400096.1	Cytoplasmic(HI_0552)	CM	CP	Virulent
WP_002289292.1	Cytoplasmic(tnp)	CM/Cytoplasm	CP	Virulent
WP_106018950.1	IM	CM/Cytoplasm	CP	Non-Virulent
WP_002288350.1	Cytoplasmic	CM	MM	Virulent
WP_106018980.1	Cytoplasmic	Cytoplasm	IS	Virulent
WP_002286553.1	Cytoplasmic	CM/Extracell	MM	Non-Virulent
WP_002303824.1	Cytoplasmic	Cytoplasm	CP	Virulent
WP_002302159.1	Extracellular	Cytoplasm	IS	Non-Virulent
WP_002346986.1	Cytoplasmic	Extracell	CP	Non-Virulent
WP_002347494.1	Cytoplasmic	Extracell	MM	Virulent
WP_002296429.1	Cytoplasmic	CM	CP	Virulent
WP_002326819.1	Cytoplasmic	CM	CP	Virulent
WP_002289272.1	Cytoplasmic	Extracell	CP	Non-Virulent
WP_002298929.1	Extracellular(ydeJ)	Cytoplasm	IS	Virulent
WP_002325767.1	Cytoplasmic	CM/Extracell	MM	Non-Virulent
WP_002350626.1	IM	Cytoplasm	CP	Virulent
WP_002326255.1	OM	CM	CP	Virulent
WP_002341874.1	OM	Extracell	MM	Virulent
WP_002289257.1	IM	Extracell	CP	Non-virulent
WP_074400044.1	Cytoplasmic	CM	MM	Virulent
WP_002286840.1	OM	CM	CP	Virulent
WP_002302142.1	Cytoplasmic	CM	MM	Virulent
WP_000185761.1	Cytoplasmic	CM	CP	Virulent
WP_106018962.1	Cytoplasmic	CM	CP	Virulent
WP_000675717.1	Periplasmic	Extracell	CP	Virulent
WP_002304624.1	OM	Extracell	CP	Virulent
WP_106018971.1	OM	Extracell	MM	Virulent
WP_002326259.1	Cytoplasmic	CM/Extracell	CP	Virulent
WP_010730972.1	Cytoplasmic(pXO2-05)	Extracell	CP	Virulent
WP_106018956.1	IM(TM_0562.1)	CM/Cytoplasm	MM	Virulent
WP_002287053.1	Cytoplasmic(pXO2-05)	CM	MM	Virulent
WP_106018969.1	Cytoplasmic	Cytoplasm/Extracell	CP	Virulent
WP_002340450.1	Cytoplasmic	CM/Cytoplasm	MM	Virulent
WP_106018948.1	Extracellular	CM	CP	Virulent
WP_002321275.1	Extracellular	Extracell	MM	Virulent
WP_002286311.1	Cytoplasmic(HI_0787)	Extracell	CP	Virulent
WP_002342384.1	Cytoplasmic	Extracell	MM	Virulent
WP_002287480.1	IM	CM	CP	Non-Virulent
WP_002317291.1	OM(yycI)	CM	CP	Virulent
WP_002288853.1	Cytoplasmic	Extracell	VF	Virulent
WP_002286524.1	Cytoplasmic	CM	CP	Virulent
WP_010730973.1	Cytoplasmic(rfaS)	CM	CP	Non-Virulent
WP_002302096.1	Cytoplasmic	CM	CP	Virulent
WP_002295906.1	Cytoplasmic	Cytoplasm	IS	Virulent
WP_002323140.1	Membrane	CIM/Cytoplasm	CP	Virulent

WP_059355966.1	Cytoplasmic	CM.	VF	Virulent
WP_106018974.1	Cytoplasmic	CM/Extracell	MM	Non-Virulent
WP_002317282.1	Cytoplasmic	CM/Extracell	VF	Non-Virulent
WP_002295913.1	Membrane	CM	MM	Virulent
WP_002321168.1	Cytoplasmic	CM	CP	Virulent
WP_000455809.1	Cytoplasmic(ypbB)	CM	CP	Virulent
WP_002305295.1	Extracellular	CM	MM	Virulent
WP_002347500.1	Cytoplasmic(aroB)	Extracell	MM	Non-Virulent
WP_002294410.1	Cytoplasmic	CM	CP	Virulent
WP_002295905.1	Cytoplasmic(yteR,yesR)	Extracell	CP	Non-virulent
WP_002286016.1	Cytoplasmic	Cytoplasm	MM	Non-virulent
WP_002288760.1	Cytoplasmic	Extracell	MM	Virulent
WP_010730971.1	Cytoplasmic(ylbC)	CM/Extracell	MM	Virulent
WP_002295486.1	Extracellular	Extracell	MM	Virulent
WP_002340448.1	CytoplasmictarF	CM	CP	Virulent
WP_002335549.1	Cytoplasmic	CM	IS	Virulent
WP_002296632.1	Cytoplasmic(ybbR)	CM	IS	Virulent
WP_002287822.1	Cytoplasmic(ydjG)	CM	CP	Non-Virulent
WP_002289233.1	Membrane	CM/Extracell	MM	Virulent
WP_002299302.1	Cytoplasmic	CM	CP	Virulent
WP_002350689.1	Cytoplasmic	Extracell	IS	Virulent
WP_002287080.1	Cytoplasmic	CM/Cytoplasm	CP	Non-Virulent
WP_049143547.1	Cytoplasmic	CM	IS	Virulent
WP_002344946.1	Extracellular	CM	CP	Virulent
WP_002288852.1	Membrane	CM/Extracell	CP	Virulent
WP_002350624.1	Cytoplasmic	CM	MM	Virulent
WP_002288487.1	Membrane	CM	IS	Non-Virulent
WP_002296549.1	Membrane	CM	MM	Non-virulent
WP_002325481.1	Extracellular	CM	MM	Virulent
WP_002287397.1	Membrane(rfbX)	Extracell	MM	Non-virulent
WP_002335550.1	Cytoplasmic	CM	MM	Virulent
WP_002341445.1	Cytoplasmic	CM	CP	Virulent
WP_002287057.1	Cytoplasmic	CM	MM	Virulent
WP_002287140.1	Cytoplasmic(SP_1800)	CM	CP	Virulent
WP_002311258.1	Cytoplasmic(yvlB)	CM	CP	Virulent
WP_002289495.1	Cytoplasmic	CM	VF	Virulent
WP_002347493.1	Cytoplasmic	Extracell	CP	Virulent
WP_002350625.1	Membrane	CM	CP	Non-Virulent
WP_002322202.1	Membrane	CM	CP	Non-Virulent
WP_002289445.1	Membrane	CM	CP	Non-Virulent
WP_002299301.1	Cytoplasmic	CM	IS	Virulent
WP_000163792.1	Extracellular(cry22Aa)	CM/Extracell	IS	Virulent
WP_002345015.1	Membrane	Extracell	CP	Non-Virulent
WP_002296543.1	Cytoplasmic	CM	CP	Virulent
WP_002354430.1	Extracellular	CM	CP	Virulent

WP_002287056.1	Extracellular	Extracell	MM	Virulent
WP_002286495.1	Cytoplasmic(helD)	CM/Extracell	VF	Non-virulent
WP_002296832.1	Extracellular(flgJ)	Extracell	MM	Virulent
WP_002325891.1	Extracellular	Extracell	CP	Virulent
WP_002296595.1	Extracellular	Extracell	VF	Virulent
WP_002296469.1	Membrane(pXO2-14)	CM/Extracell	CP	Virulent
WP_101706209.1	Extracellular	CW/Extracell	VF	Virulent
WP_002340339.1	Membrane(pXO2-14)	Extracell	VF	Virulent
WP_002340550.1	Cytoplasmic(yhaN)	Extracell	VF	Virulent
WP_002286213.1	Cytoplasmic	CIM	VF	Virulent

*CyM-Cytoplasmic membrane; CIM-Cell inner membrane; VF-Virulence Factors; CW-Cell Wall; CM-Cell Membrane; CP-Cellular Process; IS-information and Storage; MM- Metabolism Molecule;

$CELLO2GO-GO represented within the brackets ()