Proteome Exploration of *Legionella pneumophila* for Identifying Novel Therapuetics: A Hierarchical Subtractive Genomics and Reverse Vaccinology Approach

Authors:

Md Tahsin Khan[1], Araf Mahmud[1], Mahmudul Hasan[2], Kazi Faizul Azim[3], Musammath Kulsuma Begum[1], Arzuba Akter[4], Shakhinur Islam Mondal[1]*

Affiliations:

[1]*Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh*

[2]*Department of Pharmaceuticals and Industrial Biotechnologuy, Sylhet Agricultural University, Sylhet-3100, Bangladesh*

[3]*Department of Microbial Biotechnology, Sylhet Agricultural University, Sylhet-3100, Bangladesh*

[4]*Department of Biochemistry and Molecular Biology, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh*

***Corresponding author**

Dr Md Shakhinur Islam Mondal

Associate Professor

Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh

Email: shakhin200-gen@sust.edu

Phone: +353-894440254

Abstract

Legionella pneumophila, the causative agent of a serious type of pneumonia (lung infection) called Legionnaires' disease. It is emerging as an antibacterial resistant strain day by day. Hence, identification of novel drug targets and vaccine candidates is essential to fight against this pathogen. Herein attempts were taken through subtractive genomics approach on complete proteome of *L. pneumophila* to address the challenges of multidrug resistance. A total 2930 proteins from *L. pneumophila* proteome were investigated through diverse subtractive proteomics approaches, e.g., identification of human non homologous and pathogen specific essential proteins, druggability and 'anti-target' analysis, prediction of subcellular localization, human microbiome non-homology screening, protein-protein interactions studies in order to find out effective drug and vaccine targets. Only 3 were identified that fulfilled all these criteria and proposed as novel drug targets against *L. pneumophila*. Furthermore, outer membrane protein TolB was identified as potential vaccine target with better antigenicity score and allowed for further *in silico* analysis to design a unique multiepitope subunit vaccine against it. Antigenicity and transmembrane topology screening, allergenicity and toxicity assessment, population coverage analysis and molecular docking approach were adopted to generate the most potent epitopes. The final vaccine was constructed by the combination of highly immunogenic epitopes along with suitable adjuvant and linkers. The designed vaccine construct showed higher binding interaction with different MHC molecules and human immune TLR2 receptor with minimum deformability at molecular level. The translational potency and microbial expression of the vaccine protein was also analyzed using pET28a(+) vector. The present study aids in the development of novel therapeutics and vaccine candidates for efficient treatment of the infections caused by *Legionella pneumophila*. However, further wet lab based investigations and in vivo trials are highly recommended to experimentally validate our prediction.

Keywords: Legionnaires' disease; *Legionella pneumophila;* Subtractive genomics; Reverse vaccinology; Protein-protein docking; Molecular dynamics; *In silico* cloning

Introduction

Legionella pneumosphila is a human pathogen that is distributed worldwide within fresh water and biofilms. [1]. Inhalation of *Legionella spp.* cause Legionnaires' disease (LD), a severe form of pneumonia and respiratory tract infection [2] and about 90% are caused by *L. pneumophila* species [3]. *L. pneumophila* belongs to the responsibility of maximum incidences regarding water-associated diseases in the United State [4,5] and caused approximately half of all reported waterborne disease outbreaks in 2005–2006 [6,7]. *L. pneumophila* was declared as an important pathogen by the US Environmental Protection Agency (EPA) Candidate Contaminant List (CLL) due to the prevalence and seriousness of its effects on disease causing [8]. In case of disease progression, Icm/Dot type IVb secretion systems are used by *Legionella spp* in order to translocate a large repertoire of effector proteins inside the host cell cytosol [9-13], which modulate the fate of phagocytic vacuoles by preventing the fusion of phagosome-lysosome along with vacuole acidification, and by recruiting vesicles that delegate on it the properties of the endoplasmic reticulum [14]. Bacteria then damage, even completely destroy the host cell after the proliferation in this compartment.

The emergence of resistance against antibiotics or combination of antibiotics of various infectious agents poses global threats in human health [15]. At present, the majority of pathogens causing infectious diseases are resistant to more than one drugs [16]. Recent epidemiological studies revealed that *L. pneumophila* strains showed a high prevalence of resistance ranging from 50% to 100% against commonly used antibiotics. These include ceftriaxone, rifampicin, azithromycin, tigecycline, erythromycin, moxifloxacin, ciprofloxacin, doxycycline and clarithromycin and levofloxacin [17,18]. This emphasizes the urgency of developing new therapeutic antibacterial agents directing towards novel targets [15,19]. Therefore, it is exigent to identify novel therapeutic targets of *L. pneumophila.*

Identification of novel drug targets and vaccine candidates leads to drug and therapeutics component designing. The disease-based approach of the traditional method of drug discovery is both expensive, and time-consuming since significant time and devoted researchers are required to identify potential ligands. In recent years, computational aided drug targets discovery methods reduce consumption of time by eliminating compounds in this process that have limited chance of success [20,21,22,23]. Among various computational methods and strategies, subtractive genomic analysis can be regarded worthy since data regarding genome and proteome are available in various online based databases. Subtractive and comparative microbial genomics are being used for the identification of targets in a number of human pathogens like *M. tuberculosis* [24], *Burkholderia pseudomalleii* [25], *Helicobacter pylori* [26] *Pseudomonas aeruginosa* [27], *Neisseria gonorrhea* [28] and *Salmonella typhi* [29]. In case of targeting genes or proteins for drug target through subtractive and comparative microbial genome analysis, the main theme is to find such targets (gene/protein) that are essential for the pathogen and possess no homology counterpart in the host [30]. Along with subtractive genomics approach, reverse vaccinology strategy optimizes the prediction and development of novel drug and vaccine targets, especially for microorganisms that are difficult to grow in the laboratory, such as intracellular bacteria including *L. pneumophila*. Conventional vaccines may take more than 15 years to develop relying on adequate antigen expression from in vitro culture models including undesirable consequences while epitope-based vaccine prediction could be a faster approach targeting immunogenic protein through entire bacterial or viral proteome [31-33]. Reverse vaccinology approach usually identified a protein vaccine candidate using defined features, such as protein subcellular localization, topology, adhesion/antigenicity probability, epitopes, and its binding to the major histocompatibility complex (MHC) class I and II molecule [34]. Reverse vaccinology approach has been proven to prioritize and design vaccine targets against multiple pathogens [35-39].

In this study, we took an in-depth subtractive genomics and reverse vaccinology approach to identify novel therapeutic drug and vaccine target in *L. pneumophila* subsp. *pneumophila (strain Philadelphia 1 / ATCC 33152 / DSM 7513).* We particularly considered the key essential or survival proteins of the pathogen which are non-homologous to the host as well as host microbiota and screened for outer membrane proteins (OMP) and B/T-cell epitopes. At the same time, we performed the human microbiome non-homologous analysis. The highest scoring

OMP's and epitopes will facilitate future in vitro and in vivo tests for the production of drugs and suitable vaccine against intracellular *L. pneumophila* infection.

Materials and methods

The genome wide proteome exploration of *L. pneumophila* was employed for identifying novel drug and vaccine targets through subtractive genomics and reverse vaccinology approach. The overall workflow has been illustrated in Fig. 1 and Fig. 2.

Proteome retrieval and Finding paralogous sequences

The entire proteome of *L. pneumophila* subsp. *pneumophila (strain Philadelphia 1 / ATCC 33152 / DSM 7513)* (UniProt ID: UP000000609), was retrieved from UniProt(http://www.uniprot.org/proteomes/) containing total 2930 proteins [40] which were subjected to CD-Hit analysis (http://weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit) [41]. Sequence identity cutoff score was set 0.6 to eliminate redundant sequences through this process with more than 60% identity. Only non-paralogous protein sequences were transferred to further investigation excluding the redundant sequences (paralogous sequences).

Identification of human non-homologous proteins

The goal of the non-homology analysis of *L. pneumophila subsp. pneumophila* is to identify pathogen specific-proteins which are non-homologous to the human. BLASTp analysis was carried out on non-paralogous proteins against the reference proteome of *Homo sapiens* using threshold expectation value (E value) 10^{-3} through Ensembl (https://uswest.ensembl.org/Multi/Tools/Blast?db=core) [42]. Proteins which showed significant

hits assuming that they have a certain level of similarity with the human genome were filtered out leaving only the nonhomologous sequences. These nonhomologous sequences were screened. The significances of this step are to avoid any functional similarity with human proteome and to minimize any undesirable cross-reactivity of the drugs thereby to prevent binding of drugs to the active site of human homologous proteins [43].

Identification of essential nonhomologous proteins

Nonhomologous proteins were subjected to Database of Essential Genes (DEG) [44]. DEG 15.2 contains 53,885 essential genes and 786 essential non-coding sequences. BLASTp of previously selected nonhomologous proteins was carried out by selecting all the organisms present in DEG 15.2 server and using threshold expectation value 10^{-10}, a minimum bit-score cut-off of 100 as parameters, but the hits with e-value $\leq 10^{-100}$, identity $\geq$ 25%, and same annotated function of the query was selected as essential proteins. These essential proteins were employed for further study.Proteins of essential genes in a bacterium constitute a minimal genome, forming a set of functional modules, which play key roles in the emerging field, synthetic biology [44].

Metabolic pathway analysis and family prediction of hypothetical proteins

Kyoto Encyclopedia of Genes and Genomes [KEGG] contains complete metabolic pathways of the living organisms [45]. Using three letter KEGG organism code hsa and lpn of human and *Legionella pneumophila subsp. Pneumophila*respectively, all the metabolic pathways present in the host (*H. sapiens*) and the pathogen were collected separately. A manual comparison was made to identify the metabolic pathways that were only present in the pathogen, as unique to *Legionella pneumophila* subsp. *pneumophila* using KEGG PATHWAY Database [46] whereas the remaining pathways were grouped as common.The predicted human nonhomologous, essential proteins of *Legionella pneumophila subsp.pneumophila* were screened by BLASTp through KAAS [47] server at KEGG for the identification of potential drug and vaccine targets. Functional annotation of genes is provided by KAAS (KEGG Automatic Annotation Server) through BLAST comparisons against the manually curated KEGG GENES database and

metabolic proteins are specified through KO (KEGG Orthology) assignments. It automatically generates KEGG pathways that indicate those metabolic proteins. Proteins involved in these unique metabolic pathways of the pathogen and proteins assigned KO but not involved in common pathways were listed for the next step of subcellular localization analysis, excluding those proteins involved in common pathways of human and pathogen. SVMProt is a server that utilizes Support Vector Machine to classify a protein sequence into functional class from its primary sequence which includes all major classes of enzymes, transporters, receptors, channels, RNA- binding and DNA- binding proteins. TheSVMProt server [48] (http://bidd2.nus.edu.sg/cgi-bin/svmprot/svmprot.cgi) was used in order to predict the functional classes of all uncharacterized, hypothetical proteins of unique metabolic proteins of pathogen.

Prediction of subcellular localization

For the reasons of gram-negativity, *Legionella pneumophila subsp. pneumophila* proteins can be identified in five feasible subcellular locations such as cytoplasm, inner membrane, periplasm, outer membrane, and extracellular.Cytoplasmic proteins can be used as drug targets while surface membrane proteins can be used both as drug and vaccine targets [49]. To predict the localization of shortlisted non-homologous, essential proteins of the pathogen were passed through PSORTb v3.0.2 [50](http://www.psort.org/psortb/index.html), CELLO v.2.5 [51] (http://cello.life.nctu.edu.tw/), ngLOC (http://genome.unmc.edu/ngLOC/index.html). The best score for a location was counted, generated by these servers. Final location of these proteins was taken through three steps: 1) if all three servers predicted the same location for a protein, that location was selected as final result, 2) if any two of these three servers predict the same location for a protein, that location was selected as final result, 3) if they predicted three different locations for a protein then PSLpred server [52] (http://crdd.osdd.net/raghava/pslpred/) was used and the predicted location for that protein done through PSLpred server was matched with previous results predicted by PSORTb v3.0.2, CELLO v.2.5, and ngLOC servers.

Druggability Screening and antigenicity analysis

Database of DrugBank 5.1.0 contains entries including 2556 approved small molecule drugs, 962 approved biotech drugs, 112 nutraceuticalsand over 5125 experimental drugs. Additionally, 5117 non-redundant protein sequences (i.e. drug target/enzyme/carrier/transporter) are linked to these drug entries. A 'druggable' target needs to have the potentiality to bind to the drugs and drug-like molecules with high affinity. Shortlisted cytoplasmic and membrane proteins were screened through the database of DrugBank 5.1.0[53] using default parameters. Presence of drug targets from the shortlisted cytoplasmic and membrane proteins in DrugBank database indicating same biological functions can act as an evidence for their druggable properties and were grouped as current therapeutic targets while their absence indicated the novelty of those drug targets, hence were classified as 'novel therapeutic targets'. Total 45 proteins that showed no similarity after passed through DrugBank database were listed as novel drug targets and vaccine candidates. VaxiJen v2.0 [54] was used for predicting the protective antigens and subunit vaccines. Only membrane protein sequences of novel drug targets of *Legionella pneumophila* subsp. *Pneumophila* were subjected to VaxiJen v2.0 selecting threshold value 0.4. Proteins that showed antigenicity prediction >0.4 in VaxiJen v2.0 were transferred for next step as potential vaccine candidates.

'Anti-target' analysis of the essential, nonhomologous and novel drug targets

In human, anti-targets or alternative drug target towards host protein candidates include the human Ether-a-go-goR elated Gene (hERG), constitutive androstane receptor (CAR), the pregene X receptor (PXR) and P-glycoproteins (P-gp). Some membrane receptorse.g. adrenergicα_{1a} , dopaminergic D2, the muscarinic M1, and serotonergic 5-HT_{2C} are also included as 'anti-targets'. A total of 210 'anti-targets' of human were reported in the literature (Supplementary file S1) [11] and corresponding sequences of these proteins were fetched from the NCBI Protein database. BLASTp was carried out for all nonhomologous, essential 'novel drug target' proteins against these 'anti-targets' setting an E-value threshold <0.005, query length >30%, identity <30% as parameters. Nonhomologous proteins showing <30% identity against these 'anti-target' proteins were transferred to next step.

Human microbiome non-homology analysis

The correlation between gut flora and human is not merely commensal but rather is symbiotic, mutualistic relationship [59], and different beneficial functions by human microbiome were also been reported [60,61].Unintentional blocking or accidental inhibition of proteins present in thismicroflora may lead to adverse effects in the host [62]. Screening of nonhomologous, essential proteins selected as vaccine candidates and novel drug targets were subjected to BLASTp through NCBI protein blast server (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using an E-value 'cutoff score 1' against the dataset present in Human Microbiome Project server (https://www.hmpdacc.org/hmp/) "43021 (BioProject)" [63] was mentioned in the Entrez Query field while selecting Non-redundant protein sequences (nr) as the database. The Human Microbiome Project (HMP) collected microbial strains from a disease-free, healthy adult population of 239 individuals, which included 18 body habitats in five areas (oral, skin, nasal, gut, and urogenital), generating 5026 microbial community compositions [64]. Proteins showing <48% similarity were selected and transferred to next step as novel therapeutic targets and vaccine candidates. This screening analysis results assured the chance of avoiding accidental inhibition and unintentional blocking of human microflora proteins.

Protein-Protein Interaction studies

The protein-protein interaction studies (PPIs) of selected shortlisted proteins were predicted using STRING v10.5 [65]. The database identifies both physical (direct) and functional (indirect) interactions. PPIs with high confidence score (≥70%) were considered to avoid false positive results. Inhibition of essential proteins can hamper other proteins to perform correctly. Therefore, analysisof PPIs of the shortlisted proteins with others of the same strain can help to identify the best therapeutic targets. Proteins showing close interactions with at least three others were selected for further studies. Only outer membrane protein with antigenicity score >0.6 was transferred to the next stage for reverse vaccinologyafter homology modeling and validation step and the rests were listed as probable novel drug targets.

Homology modeling, structure optimization and validation

Since no exact PDB structure was available for the selected proteins, a BLASTp was carried out in NCBI for each protein where Protein Data Bank Proteins (pdb) was selected as the Database. The templates were chosen for homology modeling considering sequence identity ≥85 and query cover ≥30. EasyModeller 4.0 software was used to generate the best 3D structure of each selected therapeutic proteins[66]. The best models were subjected to ModRefiner for energy minimization and structure refinement [67]. Root mean square deviation (RMSD) values were also calculated. Evaluation of the optimized, refined structures was carried through online servers, PROCHECK [68] and ERRAT [69] present in Structural Analysis and Verification Server (SAVES) (http://servicesn.mbi.ucla.edu/SAVES/) to get the stereochemical quality of the model.

Antigenic protein selection and T-Cell epitope prediction

Among the membrane associated proteins, the final vaccine target was selected according to protein-protein interaction analysis and antigenic scoring. Different physicochemical properties of the proteins were predicted using ProtParam [70], one of the ExPASy's servers for primary structure prediction of proteins. MHC-I prediction tool (http://tools.iedb.org/mhci/) and MHC-II prediction tool (http://tools.iedb.org/mhcii/) from IEDB (Immune Epitope Database) were used to predict the MHC-I binding and MHC-II binding respectively [71].

Transmembrane topology, antigenicity, population coverage andallergenicity assessment of T-Cell epitope

The TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/) predicted the transmembrane helices in protein [72]. Again, VaxiJen v2.0 server was used to determine the epitope antigenicity [54]. In this study, population coverage for each individual epitope was analyzed by the Immune Epitope Database and Analysis Resource (IEDB) (http://tools.iedb.org/population/). Four servers named AllerTOP (http://www.ddg-pharmfac.net/AllerTop/) [73], AllergenFP (http://www.ddg-pharmfac.net/AllergenFP/) [74], PA3P

(http://www.Ipa.saogabriel.unipampa.edu.br:8080/pa3p/) [75] and Allermatch (http:// www.allermatch.org/allermatch.py/form) [76] were used to predict the allergenicity of the proposed epitopes for vaccine development.

Toxicity and conservancy analysis of T-Cell epitope

The relative toxicity level was demonstrated by ToxinPred server (http://crdd.osdd.net/raghava/toxinpred/).Epitope conservancy is a crucial step in the immunoinformatic approach as it determines the degree of desired epitope distributions in the homologous protein set. IEDB's epitope conservancy analysis tool (http://tools.iedb.org/conservancy/) was selected for the analysis of conservancy level by concentrating on the identities of the selected protein.

B-Cell epitope identification and epitope cluster analysis

B cell epitope prediction tools from IEDB were used to identify the B cell antigenicity based on six different algorithms which include Kolaskar and Tongaonkar antigenicity scale [77], Emini surface accessibility prediction [78],Karplus and Schulz flexibility prediction [79], Bepipred linear epitope prediction analysis [80], Chou and Fasman beta turn prediction [81] and Parker hydrophilicity prediction [82].To find out the overlapping peptides, IEDB epitope cluster analysis resource was utilized [83].

Final vaccine construction

Larger cassettes containing multiple epitopes were identifiedamong top CTL, HTL and BCL epitopes.The predicted T-cell and B-cell epitopes were conjugated in a sequential manner to design the final vaccine construct. All three vaccine proteins started with an adjuvant followed by the top CTL epitopes for the selected protein, then by top HTL epitopes and BCL epitopes

respectively, in the similar fashion. Three vaccine sequence were constructed named V1, V2 and V3, each associated with different adjuvants including beta defensin (a 45 mer peptide), L7/L12 ribosomal protein and HABA protein (*M. tuberculosis*, accession number: AGV15514.1). Interactions of adjuvants with toll like receptors (TLRs) polarize CTL responses and induce robust immunoreactions [84]. Beta defensin adjuvant can act as an agonist to TLR1, TLR2 and TLR4 receptor. On the contrary, L7/L12 ribosomal protein and HBHA protein are agonists to TLR4 only. To overcome the problems caused by highly polymorphic HLA alleles, PADRE sequence was also incorporated along with the adjuvant peptides. EAAAK linkers were used to join the adjuvant and CTL epitopes. Similarly, GGGS,GPGPG and KK linkers were used to conjugate the CTL, HTL and BCL epitopes respectively. Utilized linkers ensured the effective separation of individual epitopes in vivo [85,86].

Allergenicity, antigenicity prediction, physicochemical characterization and secondary structure analysis of vaccine protein

AlgPred v.2.0 [87] sever was used to predict the non-allergic nature of the constructed vaccines. In order to suggest the superior vaccine candidate, VaxiJen v2.0 server [54] was used to evaluatethe probable antigenicity of the vaccine constructs. ProtParam, a tool provided by Expasy server [70,88] functionally characterized the vaccine constructs according to molecular weight, aliphatic index, isoelectric pH, hydropathicity, instability index, GRAVY values, estimated half-life and various physicochemical properties. The PSIPRED v3.3 [89] and NetTurnP 1.0 program [90,91] was used to predict the alpha helix, beta sheet and coil structure of the vaccine constructs.

Vaccine tertiary structure prediction, disulfide engineering and protein-protein docking

The RaptorX server performed 3D modeling of the designed vaccines depending on the degree of similarity between target protein and available template structure from PDB [92,93]. Refinement was conducted using ModRefiner [94] followed by FG-MD refinement server [95] to improve the accuracy of the predicted 3D modeled structure. The refined protein structure was

further validated by Ramachandran plot assessment at RAMPAGE [96].Disulfide bonds enhance the geometric conformation of proteins and provide significant stability. DbD2, an online tool was used to design such bonds for the constructed vaccine protein [97]. Inflammations caused by bacterial antigen are involved with TLR-2 immune receptors present over the immune cells [98-100]. An approach for protein-protein docking was employed to determine the binding affinity of designed subunit vaccines with different HLA alleles and TLR-2 immune receptor by using ClusPro [101], hdoc [102,103] and PatchDock server [104]. FireDockserver refined the complexes generated via PatchDock server [104]. The 3D structure of different MHC molecules and human TLR-2 receptor was retrieved from RCSB protein data bank.

Molecular dynamics simulation

Stability can be determined by comparing the essential dynamics of proteins to their normal modes [105,106].This powerful tool is an alternative to the costly atomistic simulation [106,107]. iMODS server explains the collective motion of proteins by analyzing the normal modes (NMA) in internal coordinates [108,109]. The structural dynamics of protein complex was investigated by using this server due to its much faster and effective assessments than other molecular dynamics (MD) simulations tools [110,111].The direction and extent of the immanent motions of the complex was analyzed in terms of deformability, eigenvalues, B-factors and covariance [112].

Codon adaptation and in silico cloning

Due to the lack of similarities between the codon usage of human and *E. coli*, codon adaptation approachwas adopted to achieve higher expression of vaccine protein V1 in the selected host. Rho independent transcription termination, prokaryote ribosome-binding site and cleavage sites of several restriction enzymes (i.e. HindIII and BamHI) were avoided during the operation performed by JCAT server [113]. The optimized sequence of vaccine protein V1 was reversed and then conjugated with HindIII and BamHI restriction site at the N-terminal and C-terminal

sites respectively. SnapGene restriction cloning module was used to insert the adapted sequence between HindIII (173) and BamHI (198) of pET28a(+) vector [114].

Results

All the methodologies used in the proteome exploration of *Legionella pneumophila* including the number of proteins screened in each step had been summarized in Table 1.

Proteome retrieval and Finding paralogous sequences

The whole proteome retrieved from UniProt (UniProtID: UP000000609) containing 2930 proteins were analyzed in CD-Hit tool using threshold value 0.6 to eliminate paralogous and duplicate sequences showing >60% sequence identity. Paralogous protein sequences were removed leaving total 2878 non-paralogous protein sequences.

Identification of human non-homologous proteins

Non-paralogous protein sequences were screened by BLASTp through Ensembl genome browser 92. 868 proteins was found showing similarity with the human genome which were excluded leaving total 2010 non-homologous proteins.

Identification of essential proteins

Among 2010 nonhomologous proteins, only 125 protein sequences showed similarity with the essential proteins enlisted in DEG server from 48 bacterial strains. These 125 proteins were listed as essential proteins and are considered as responsible for the survival of *Legionella pneumophila*.

Metabolic pathway analysis and family prediction of hypothetical proteins

Though 120 pathways of target *Legionella pneumophila* were identified from KEGG server, only 37 pathways were detected as unique for pathogen (Supplementary file S2). The result from KAAS server at KEGG revealed that 122 proteins among 125 had assigned KO identifier. From 122, only 4 proteins were involved in unique metabolic pathways (Table 2) and 29 proteins involved neither in unique nor common pathways (Supplementary Table 1). Only 3 hypothetical proteins were found among these 29 KO assigned proteins for which functions remains unknown. Functional family for these hypothetical proteins were also analyzed (Supplementary Table 2). All of these 33 proteins were allowed for further investigation. Remaining 89 proteins were excluded since these were found to be involved in both pathogen and human metabolic pathways.

Prediction of subcellular localization

The shortlisted 33 nonhomologous essential proteins involved in unique pathways of the pathogen were screened through PSORTb, CELLO and ngLOC servers. A total 16 proteins were identified as cytoplasmic proteins and 16 proteins were identified as membrane proteins which were further screened for druggability and antigenicity analysis (Supplementary file S3)**.** But, the study failed to predict the subcellular localization of 'tr|Q5ZYU2' since these four servers generated different predictions.

Druggability Screening and antigenicity analysis

Only 7 proteins showed similarity with the available drug targets (Table 3) at DrugBank5.1.0 database. The remaining 25 showed no hitand thesewere considered as novel drug targets of *L. pneumophila* which were comprised of 13 cytoplasmic and 12 membrane proteins. Antigenicity analysis of membrane protein revealed that only 1 protein was non antigenic whereas the rest 11 showed antigenicity score more than 0.4. at VaxiJen server (Table 4). Furthermore, all of the

proteins screened after druggability and antigenicity analysis were employed for human 'anti-targets' and microbiome non-homology study.

'Anti-target' and Human microbiome non-homology analysis

To avoid severe cross-reaction and toxic effects in human, identification of nonhomologous proteins to human essential proteins (referred as 'anti-targets') is a crucial step. Study revealed that all selected 25novel therapeutic proteins showed no similarity with 210 human 'anti-targets'. In addition, BLASTp against all microbial strains presented in Human Microbiome Project (HMP) server carried out through NCBI blast server revealed that only 5 proteins out of 25 showed similarity ≤47% (Table 5). Targeting these proteins will be suitable since they are neither involved in common pathways of host-pathogen nor homologous to any human 'anti-targets'.

Protein-Protein Interaction studies

PPIs in STRING v10.5 revealed that Cytoplasmic protein 'UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase (murE)' showed interactions with proteins involved in cell wall biosynthesis pathways (Fig. 3A) whereas 'Trigger factor (tig)' showed interactions with proteins which were mainly involved in binding with different rRNA and tRNA molecules (Fig. 3B). Moreover, inner membrane protein 'Probable lipid II flippaseMurJ (mviN)' exhibited interactions with proteins which were connected to peptidoglycan biosynthesis pathway and transportation of lipid-linked peptidoglycan precursors (Fig. 3C), where 'Biotin synthase (bioB)' is an essential protein for the survival of pathogen. The name of essential drug targets listed in Table 6. Another inner membrane protein 'Probable potassium transport system protein kup 1 (kup1)' presented single interaction (Fig. 3D) and as a consequence it was excluded from our analysis. Moreover, the remaining outer membrane protein 'TolB (tolB)' showed close interactions with different other proteins (Fig. 3E) which were mainly found in peptidoglycan synthesis and lipoproteins translocation process. 'Protein TolB' also exhibited less similarity (<47%) with human microflora proteins and had antigenicity score >0.6, that is why, it had been

considered as potential vaccine candidate against *Legionella pneumophila*. So, further study was employed for 'in silico vaccine design' focusing 'Protein TolB' by reverse vaccinology approach. Details of PPIs was provided in Supplementary file S4.

Homology modeling, structure optimization and validation

The selected single vaccine candidate Q5ZV69 was modeled to determine 3D structure through EasyModeller 4.0 (Supplementary Fig. 1A). Complete results of template selection for homology modeling are provided in Supplementary Table 3. The best structure generated for each target was optimized by ModRefiner server. Refined model of Q5ZV69 had RMSD 0.1444 and TM score 0.9783 to initial model. After refinement, the PDB structure of each protein was evaluated through PROCHECK and ERRAT. Psi-Phi pains in Ramachandran Plot revealed that protein Q5ZV69 had 91.9% residues in most favored regions, 7.8% residues in additional allowed region and 0.3% in generously allowed region (Supplementary Fig. 1B). According to ERRAT, the overall quality factor of this protein was 76.1084 (Supplementary Fig. 2). An ERRAT score of 50 is normally acceptable and the value explains the statistics of non-bonded interaction molecules. This indicates that the modeled structure is reliable and stable.

Antigenic protein selection and T-Cell epitope prediction

From the top membrane proteins, LEGPH Protein TolB (Accession ID: Q5ZV69) was selected based on protein-protein interaction analysis with total antigenicity score of 0.6740 for vaccine design against *Legionella*. Various physiochemical parameters of the proteins were analyzed by ProtParam tools as listed in Supplementary Table 4. Further,numerous immunogenic epitopes from LEGPH Protein TolB were identified to be potential T cell epitopes using both MHC-I and MHC- II binding predictions of the IEDB that can bind a large number of different HLA-A and HLA-B alleles with high binding affinity. Epitopes that bind to the maximum number of HLA cells were selected.

Transmembrane topology, antigenicity, population coverage and allergenicity assessment of T-Cell epitope

Top epitopes of LEGPH Protein TolB were selected as putative T cell epitope candidates based on their transmembrane topology screening by TMHMM and antigenic scoring (AS) by Vaxijen (Table 7). Population coverage for each individual epitopes was analyzed by the IEDB population coverage calculation tool analysis resource. All indicated alleles were identified as optimum binders with the predicted epitopes and were used to determine the population coverage (Fig. 4). Epitopes that were found to be non-allergen for human were identified by four different allergenicity prediction servers. Epitopes, indicated as allergenic for human or classified as undefined were removed from the suggested list of epitopes (Table 7).

Toxicity and conservancy analysis of T-Cell epitope

ToxinPred server predicted the relative toxicity of the top epitopes (Table7). Again, putative epitopes generated from LEGPH Protein TolB were found to be highly conserved with maximum 100% conservancy level. Top epitopes were used to construct the final vaccine candidates to ensure a broad spectrum immune response (Table 7).

B-Cell epitope identification and epitope cluster analysis

B-cell epitopes of envelope glycoprotein and matrix protein VP40 were generated using a total six algorithms (Table 8).For envelope glycoprotein, Bepipred prediction method predicted the peptide sequences from 66-77 and 404-417 amino acids as potential B cell epitopes that could induce the preferred immune responses (Supplementary Fig. 3A). Emini-surface accessibility prediction was also conducted which indicated 206-212 and 277-287 amino acid residues to be more accessible (Supplementary Fig. 3B). Chou and Fasman beta-turn prediction method displayed regions from 194-200 and 294-300 as potential Beta-turn regions (Supplementary Fig. 3C). Karplus and Schulz flexibility prediction found the region of 160-166 and 308-314 amino acid residues as most flexible regions (Supplementary Fig. 3D). In contrast, Kolaskar and Tongaonkar antigenicity result confirmed the region from 5-23 and 243-252 as highly antigenic

(Supplementary Fig. 3E) while Parker hydrophilicity prediction indicated 174-180 and 283-289 amino acid residues to be more potent (Supplementary Fig. 3F). Finally, IEDB epitope cluster analysis resource was used to identify the clusters (peptides containing multiple epitopes) among top CTL, HTL and BCL epitopes (Supplementary Table 5).

Final vaccine construction

Both Clusters and singletons (contains single epitope) were used to design the final vaccine construct. PADRE sequence was incorporated to maximize the efficacy and potency of the peptide vaccine. A total 3 vaccines i.e. V1, V2 and V3 (476, 561, and 590 amino acid longrespectively) were constructed (Table 9) and further analyzed to investigate their immunogenic potential.

Allergenicity, antigenicity prediction, physicochemical characterization and secondary structure analysis of vaccine protein

All three constructs (V1, V2 and V3) were non-allergic in behavior. However, V1 was best in terms of safety and found superior as potential vaccine candidate with better antigenicity (0.603) and ability to stimulate preferred immune response (Table 9). The final vaccine construct was characterized on the basis of physical and chemical properties. The molecular weight of the vaccine construct V1 was 48.34kDa which ensured its good antigenic potential. The theoretical pI 9.78 indicated that the protein will have net negative charge above the pI and vice versa. The extinction coefficient was 41370, assuming all cysteine residues are reduced at 0.1% absorption. The estimated half-life of the constructed vaccine was expected to be 1h in mammalian reticulocytes in vitro while more than 10 h in *E. coli* in vivo. Thermostability and hydrophilic nature of the vaccine protein was represented by aliphatic index and GRAVY value which were 61.74 and -0.545 respectively. The computed instability index of the protein was 33.51 which classified it as a stable one. Secondary structure of the construct V1 confirmed to have 6.30% alpha helix, 24.78% sheet and 68.92% coil structure (Supplementary Fig. 4).

Vaccine tertiary structure prediction, disulfide engineering and protein-protein docking

RaptorX generated the tertiary structure of the designed construct V1 consisting single domain (Fig. 5A). Homology modeling was performed by detecting and using 2ymuA from protein data bank (PDB) as best suited template for Vaccine V1. All 476 amino acid residues (100%) were modeled with only 8% residues in the disordered region. The quality of the 3D model was defined by P value which was 5.30e-11 for the predicted vaccine protein. The low P value ensured better model quality of the predicted vaccine. Refinement was performed to improve the quality of predicted 3D modeled structure beyond the accuracy.Before refinement Ramachandran plot analysis revealed that 81.4% residues were in the favored, 11.8% residues in the allowed and 6.8% residues in the outlier region. However, after refinement 88.4% and 8.4% residues were in the favored and allowed region respectively, whileonly 3.2% residues were found in the outlier region (Fig. 5B). Modeled tertiary structure of vaccine construct V2 and V3 have been shown in Supplementary Fig. 5. Residues in the highly mobile region of the protein sequence were mutated with cysteine to perform Disulfide engineering. A total 71 pairs of amino acid residue were identified having the capability to form disulfide bond by DbD2 server. After evaluation of the residue pairs in terms of energy, chi3 and B-factor parameter, only 3 pairs () satisfied the disulfide bond formation. Those residue pairs were ASN202-GLY 226, SER 211-GLY 217 and ASN 309-LEU 317 (Fig. 6). All these 4 residues were replaced with cysteine residue. The value of chi3 considered for the residue screening was between -87 to +97 while the energy value was less than 2.5.Docking analysis was performed between vaccine construct and different HLA alleles (Table 10).Construct V1 showed biologically significant results in terms of free binding energy. Besides, docking was also conducted to evaluate the binding affinity of designed vaccine with human TLR2 receptor (Fig. 7). The ClusPro server generated 30 protein-ligand complexes as output along with respective free binding energy. The lowest energy of -1232.8 was obtained for the complex named cluster 1. The hdoc server predicted the binding energy for the protein-protein complex was -275.55, while FireDock output refinement of PatchDock server showed the lowest global energy of -15.64 for solution 2. The lowest binding energy of the complexes indicates the highest binding affinity between TLR-2 and vaccine construct.

Molecular dynamics simulation

Normal mode analysis (NMA) was performed to investigate the stability of proteins and their mobility at large scale. Deformabilty depends on the individual distortion of each residue, represented by hinges in the chain (Fig. 8A). The B-factor value was equivalent to RMS inferred via NMA (Fig. 8B). The degree of mobility was indicated by the length of the line. The eigenvalue found for the complex was 3.5007e-06 (Fig. 8C). There was an inverse relationship between eigenvalue and the variance associated to each normal mode (Fig. 8D). Coupling between pairs of residues was indicated by the covariance matrix where different pairs showed correlated, anti-correlated or uncorrelated motions, represented by red, blue and white colors respectively (Fig. 8E). An elastic network model was also generated to detect the pairs of atoms connected via springs (Fig. 8F).

Codon adaptation and in silico cloning

Due to dissimilarity in the regulatory systems of human and *E. coli*, codon adaptation was performed considering the expression system of the host. Codon adaptation index (CAI) was 0.95 ensuring the higher proportion of most abundant codons. The GC content of the optimized codons (54.01%) was significant as well. Finally, the optimized codons were inserted into pET28a(+) vector along with HindIII and BamHI restriction sites. A clone of 6346 base pair was produced comprising 1434 bp desired sequence (shown in red color in between the sequence of pET28a(+) vector) and the rest belonging to the vector only (Fig. 9).

Discussion

Legonellapneumophila is one of the main pathogenic agents responsible for pneumonia and respiratory tract infections while showing high levels of resistance against a large number of antibiotics. Only a few small, controlled clinical studies of antibiotic treatment for Legionnaires' diseasehave been completed [115]. Hence, the evidence for treatment recommendations is limited [116]. However, Rahimi and Vesalreported that *L. pneumophila*strains harboredhighest levels of resistance againstciprofloxacin, erythromycin, clarithromycin, moxifloxacin andazithromycin [117]. Therefore, the ultimate goal of this study

was to identify and determine novel drug targets and vaccine candidates for *Legionella pneumophila*through subtractive genomics and reveres vaccinology approach.

Computational predictions help to identify proteins those are essential for the survival of the pathogen, non-homologous to host proteins, as well as no involvement in metabolic pathways of host leading to choose the proteins only present and involve in the metabolic pathways of the pathogen. Complete proteome of *L. pneumophila* (2,930 proteins) were retrieved from the NCBI protein database. Proteins involved in several common cellular systems emerged as homologous between bacteria and human in course of evolution [118,119] which were removed in this study based on their identity with human proteins. Essential proteins are most promising for new antibacterial drug targets since most antibiotics are designed to bind essential gene products [120] and can be considered as species-specific drug targets [121]. The study revealed 125 unique essential proteins of *L. pneumophila* which can be considered as suitable drug targets. Localization is an important aspect of any possible drug target since cellular functions of proteins often are localized in specific compartments, hence studying subcellular localization will be helpful to understand disease mechanisms as well as developing new vaccine candidates and drug targets. Although, both membrane and cytoplasmic proteins could serve the purpose as therapeutic targets [122], membrane proteins habe been mostly reported for vaccine candidates [123]. Hence, in this study membrane proteins were used for vaccine construction whereas cytoplasmic proteins were proposed as suitable drug targets. Again, usage of antibiotics has already reduced their efficiency leading to the rapid emergence of resistant bacteria [124]. The antibiotic resistance crisis is observed due to the misuse as well as overuse of these drugs, and a scarcity of developing new drugs by the pharmaceutical industry [125,126]. Broad-spectrum drugs medicated for a pathogen or group of the pathogen may cause mutational changes as well as enhance the transfer of gene to other pathogens which can show resistance to drugs leading towards the emergence of resistant bacteria. To avoid this crisis, selection of novel drug targets and vaccine candidates is a must; therefore shortlisted 32 proteins were further screened through DrugBank5.1.0 database with default parameters.

Antigenicity score for selected 12 novel drug targeted membrane proteins revealed that only 1 protein was non antigenic whereas the rest 11 showed antigenicity score >0.4. These 11 proteins can be used to design B cell and T cell epitopes in future to lessen the disease caused by

Legionella pneumophila. A number of drugs from different categories such as non-nacrotic analgesics, non-steroidal anti-inflammatory drugs, antidepressants as well as vasodilators were withdrawn during 1960-1999 due to various health-relatedcross-reactivity causing hepatic, cardiovascular, dermatologist and hematologic toxicities as well as carcinogenic effects [127]. For example, bromfenac [128], ebrotidine [129], trovafloxacin [130] were withdrawn from pharmaceutical markets worldwide since they showed hepatotoxic effects. Therefore, specific recognition must be maintained by an ideal drug target to the drug-treated against it, furthermore, this drug target needs to be significantly different from the host proteins. To avoid severe cross-reaction and toxic effects in human, identification of nonhomologous proteins to human essential proteins (referred as 'anti-targets') was a crucial step considered in this study. Targeting Human microbiome non-homology proteins will be suitable since they are neither involved in common pathways of host-pathogen nor homologous to any human 'anti-targets'.Furthermore drug or vaccine designed and administered for these novel targets will be less harmful to other commensal microbial strains dwelling in different body sites of a healthy human. The study revealed that most of the cytoplasmic proteins are involved in cell wall biosynthesis pathways and essential for cell division. However, targeting three proteins i.e. UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase, Trigger factor and Probable lipid II flippaseMurJ also influence the interacting proteins can lead to development of novel therapeutics in future. The predicted single vaccine candidate Q5ZV69 in the present study was analyzed to design a potent highly immunogenic vaccine candidate against *Legionella pneumophila*. Numerous antigenic epitopes were generated which were investigated extensively for antigenicity, allergenicity, toxicity, conservancy and other physiochemical properties using a number of bioinfomatics tools and software. Final vaccine constructs were designed with the help of different adjuvants and amino acids linkers. It has been reported that PADRE sequence reduces the polymorphismof HLA molecules in the population [131,132]. Linkers in vaccine also enhanced the immunogenicity of the vaccines in previous studies [133,134]. Therefore, we have taken all important factors that can induce the immunogenicity of our vaccine constructs. Moreover, docking analysis was done to explore the binding affinity of promiscuous epitopes with different HLA alleles including HLA-DRB1*03:01 (1A6A), (HLA-DRB5*01:01 (1H15), HLA-DRB3*01:01 (2Q6W), HLA-DRB1*04:01 (2SEB), HLA-DRB1*01:01 (2FSE), and HLA-DRB3*02:02 (3C5J). Furthermore, interaction between vaccine

construct V1 and human immune-receptor TLR 2 were also checked to demonstrate the efficacy of used adjuvant. The vaccine-receptor complex showed deformability at a minimum level which also strengthened our prediction.

Conclusion:

By using the subtractive genomics and reverse vaccinology approach, we are trying to develop novel therapeutics against *Legionella pneumophila* and may help to reduce the mortality and morbidity caused by it. We particularly considered the key essential or survival proteins of the pathogen which are non-homologous to the host as well as host microbiota and screened for outer membrane proteins (OMP) and B/T-cell epitopes. At the same time, we performed the human microbiome non-homologous analysis. The highest scoring OMP's and epitopes will facilitate future in vitro and in vivo tests for the production of drugs and suitable vaccine against intracellular *L. pneumophila* infection. After intensive analysis, only 3 proteins were identified and proposed as novel drug targets against *L. pneumophila*. Outer membrane protein TolB was identified as potential vaccine target with better antigenicity score. However further *in vivo* trials using model organisms are highly recommended to validate our prediction.

References

1. Chien M, Morozova I, Shi S, Sheng H, Chen J, et al. (2004) The genomic sequence of the accidental pathogen Legionella pneumophila. Science 305: 1966-1968.
2. National Research Council (US) Committee on Drinking Water Contaminants. Review of Methods for Assessing Microbial Pathogens; National Academies Press (US):Washington, DC, USA, 1999.
3. Yu VL, Plouffe JF, Pastoris MC, Stout JE, Schousboe M, Widmer A, Summersgill J, File T, Heath CM, Paterson DL, Chereshsky A. Distribution of Legionella species and serogroups isolated by culture in patients with sporadic community-acquired legionellosis: an international collaborative survey. The Journal of infectious diseases. 2002 Jul 1;186(1):127-8.

4. Kaufmann, A.F.; McDade, J.E.; Patton, C.M.; Bennett, J.V.; Skaliy, P.; Feeley, J.C.; Anderson, D.C.; Potter, M.E.; Newhouse, V.F.; Gregg, M.B.; et al. Pontiac Fever: Isolation of the Etiologic Agent (Legionella pneumophila) and Demonstration of its Mode of Transmission. Am. J. Epidemiol. 1981, 114, 337–347.
5. Neil, K.; Berkelman, R. Increasing incidence of legionellosis in the United States, 1990–2005: Changing epidemiologic trends. Clin. Infect. Dis. 2008, 47, 591–599.
6. Centers for Disease Control and Prevention (CDC). Legionellosis—United States, 2000–2009. MMWR Morb. Mortal. Wkly. Rep. 2011, 60, 1083–1086.
7. Choffnes, E.R.; Mack, A. Global Issues inWater, Sanitation, and Health: Workshop Summary; National Academies Press: Washington, DC, USA, 2009.
8. Mraz A, Weir M. Knowledge to Predict Pathogens: Legionella pneumophila Lifecycle Critical Review Part I Uptake into Host Cells. Water. 2018 Jan 31;10(2):132.
9. Zamboni DS, McGrath S, Rabinovitch M, Roy CR (2003) Coxiella burnetii express type IV secretion system proteins that function similarly to components of the Legionella pneumophila Dot/Icm system. Molecular microbiology 49: 965-976.
10. Zusman T, Yerushalmi G, Segal G (2003) Functional similarities between the icm/dot pathogenesis systems of Coxiella burnetii and Legionella pneumophila. Infection and immunity 71: 3714-3723.
11. Nagai H, Kagan JC, Zhu X, Kahn RA, Roy CR (2002) A bacterial guanine nucleotide exchange factor activates ARF on Legionella phagosomes. Science 295: 679-682.
12. Luo Z-Q, Isberg RR (2004) Multiple substrates of the Legionella pneumophila Dot/Icm system identified by interbacterial protein transfer. Proceedings of the National Academy of Sciences of the United States of America 101: 841-846.
13. Chen J, de Felipe KS, Clarke M, Lu H, Anderson OR, et al. (2004) Legionella effectors that promote nonlytic release from protozoa. Science 303: 1358-1361.
14. Nagai H, Roy CR (2003) Show me the substrates: modulation of host cell function by type IV secretion systems. Cellular microbiology 5: 373-383.
15. Chopra I, Schofield C, Everett M, O'Neill A, Miller K, et al. (2008) Treatment of health-care-associated infections caused by Gram-negative bacteria: a consensus statement. The Lancet infectious diseases 8: 133-139.

16. Uddin R, Saeed K (2014) Identification and characterization of potential drug targets by subtractive genome analyses of methicillin resistant Staphylococcus aureus. Computational biology and chemistry 48: 55-63.
17. Nielsen K, Bangsborg JM, Høiby N (2000) Susceptibility of Legionella species to five antibiotics and development of resistance by exposure to erythromycin, ciprofloxacin, and rifampicin. Diagnostic microbiology and infectious disease 36: 43-48.
18. De Giglio O, Napoli C, Lovero G, Diella G, Rutigliano S, et al. (2015) Antibiotic susceptibility of Legionella pneumophila strains isolated from hospital water systems in Southern Italy. Environmental research 142: 586-590.
19. Newton HJ, Ang DK, Van Driel IR, Hartland EL. Molecular pathogenesis of infections caused by Legionella pneumophila. Clinical microbiology reviews. 2010 Apr 1;23(2):274-98.
20. Azim KF, Hasan M, Hossain MN, Somana SR, Hoque SF, Bappy MN, Chowdhury AT, Lasker T. Immunoinformatics approaches for designing a novel multi epitope peptide vaccine against human norovirus (Norwalk virus). Infection, Genetics and Evolution. 2019 Jun 21:103936.
21. Hasan M, Azim KF, Begum A, Khan NA, Shammi TS, Imran AS, Chowdhury IM, Urme SR. Vaccinomics strategy for developing a unique multi-epitope monovalent vaccine against Marburg marburgvirus. Infection, Genetics and Evolution. 2019 Jun 1;70:140-57.
22. Elliott SL, Suhrbier A, Miles JJ, Lawrence G, Pye SJ, Le TT, Rosenstengel A, Nguyen T, Allworth A, Burrows SR, Cox J. Phase I trial of a CD8+ T-cell peptide epitope-based vaccine for infectious mononucleosis. Journal of virology. 2008 Feb 1;82(3):1448-57.
23. Hosen MJ, Hasan M, Chakraborty S, Abir RA, Zubaer A, Coucke P. Comprehensive in silico Study of GLUT10: Prediction of Possible Substrate Binding Sites and Interacting Molecules. Current pharmaceutical biotechnology. 2019.
24. Asif SM, Asad A, Faizan A, Anjali MS, Arvind A, Neelesh K, et al. Dataset of potential targets for Mycobacterium tuberculosis H37Rv through comparative genome analysis. Bioinformation. 2009; 4(6):245– 8.
25. Chong CE, Lim BS, Nathan S, Mohamed R. In silico analysis of Burkholderia pseudomallei genome sequence for potential drug targets. In Silico Biol. 2006; 6(4):341–6.

26. Dutta A, Singh SK, Ghosh P, Mukherjee R, Mitter S, Bandyopadhyay D. In silico identification of potential therapeutic targets in the human pathogen Helicobacter pylori. In Silico Biol. 2006; 6(1–2):43–7.
27. Sakharkar KR, Sakharkar MK, Chow VT. A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa. In Silico Biol. 2004; 4(3):355– 60.
28. Barh D, Kumar A. In silico identification of candidate drug and vaccine targets from various pathways in Neisseria gonorrhoeae. In Silico Biol. 2009; 9(4):225–31.
29. Rathi B, Sarangi AN, Trivedi N. Genome subtraction for novel target definition in Salmonella typhi. Bioinformation. 2009; 4(4):143–50.
30. Barh D, Jain N, Tiwari S, Parida BP, D'Afonseca V, Li L, et al. A novel comparative genomics analysis for common drug and vaccine targets in Corynebacterium pseudotuberculosis and other CMN group of human pathogens. Chem Biol Drug Des. 2011; 78(1):73–84.
31. Poland GA, Ovsyannikova IG, Jacobson RM. Application of pharmacogenomics to vaccines. Pharmacogenomics 2009; 10 (5): 837-52.
32. Flower DR. Bioinformatics for vaccinology. John Wiley & Sons 2008; 1st edition: 302.
33. Wong SS, Jeevan T, Kercher L, Yoon SW, Petkova AM, Crumpton JC, et al. A single dose of whole inactivated H7N9 influenza vaccine confers protection from severe disease but not infection in ferrets. Vaccine 2014; 32: 4571-7.
34. Xiang Z, He Y. Genome-wide prediction of vaccine targets for human herpes simplex viruses using Vaxign reverse vaccinology. BMC bioinformatics. 2013 Mar;14(4):S2.
35. Meunier M, Guyard-Nicodème M, Hirchaud E, Parra A, Chemaly M, Dory D. Identification of novel vaccine candidates against Campylobacter through reverse vaccinology. Journal of immunology research. 2016;2016.
36. Mora M, Telford JL. Genome-based approaches to vaccine development. Journal of Molecular Medicine. 2010 Feb 1;88(2):143-7.
37. Barh D, Tiwari S, Jain N, Ali A, Santos AR, Misra AN, Azevedo V, Kumar A. In silico subtractive genomics for target identification in human bacterial pathogens. Drug Development Research. 2011 Mar;72(2):162-77.

38. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, et al. (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nature reviews Drug discovery 9: 203.
39. Hasan M, Ghosh PP, Azim KF, Mukta S, Abir RA, Nahar J, Khan MM. Reverse vaccinology approach to design a novel multi-epitope subunit vaccine against avian influenza A (H7N9) virus. Microbial pathogenesis. 2019 May 1;130:19-37.
40. Chien, M., et al., The genomic sequence of the accidental pathogen Legionella pneumophila. Science, 2004. 305(5692): p. 1966-1968.
41. Huang, Y., et al., CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics, 2010. 26(5): p. 680-682.
42. Zerbino, D.R., et al., Ensembl 2018. Nucleic acids research, 2017. 46(D1): p. D754-D761.
43. Shanmugham, B. and A. Pan, Identification and characterization of potential therapeutic candidates in emerging human pathogen Mycobacterium abscessus: a novel hierarchical in silico approach. PLoS One, 2013. 8(3): p. e59126.
44. Luo, H., et al., DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. Nucleic acids research, 2013. 42(D1): p. D574-D580.
45. Kanehisa, M. and S. Goto, KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research, 2000. 28(1): p. 27-30.
46. Qiu, Y.-Q., KEGG pathway database. Encyclopedia of Systems Biology, 2013: p. 1068-1069.
47. Moriya, Y., et al., KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic acids research, 2007. 35(suppl_2): p. W182-W185.
48. Li, Y.H., et al., SVM-Prot 2016: a web-server for machine learning prediction of protein functional families from sequence irrespective of similarity. PLoS One, 2016. 11(8): p. e0155290.
49. Damte, D., et al., Putative drug and vaccine target protein identification using comparative genomic analysis of KEGG annotated metabolic pathways of Mycoplasma hyopneumoniae. Genomics, 2013. 102(1): p. 47-56.

50. Yu, N.Y., et al., PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics, 2010. 26(13): p. 1608-1615.
51. Yu, C.S., C.J. Lin, and J.K. Hwang, Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein science, 2004. 13(5): p. 1402-1406.
52. Bhasin, M., A. Garg, and G. Raghava, PSLpred: prediction of subcellular localization of bacterial proteins. Bioinformatics, 2005. 21(10): p. 2522-2524.
53. Wishart, D.S., et al., DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic acids research, 2017. 46(D1): p. D1074-D1082.
54. Doytchinova, I.A. and D.R. Flower, VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC bioinformatics, 2007. 8(1): p. 4.
55. Kinross, J.M., A.W. Darzi, and J.K. Nicholson, Gut microbiome-host interactions in health and disease. Genome medicine, 2011. 3(3): p. 14.
56. Reid, G., et al., Microbiota restoration: natural and supplemented recovery of human microbial communities. Nature Reviews Microbiology, 2011. 9(1): p. 27.
57. Schommer, N.N. and R.L. Gallo, Structure and function of the human skin microbiome. Trends in microbiology, 2013. 21(12): p. 660-668.
58. Cogen, A., V. Nizet, and R. Gallo, Skin microbiota: a source of disease or defence? British Journal of Dermatology, 2008. 158(3): p. 442-455.
59. Sears, C.L., A dynamic partnership: celebrating our gut flora. Anaerobe, 2005. 11(5): p. 247-251.
60. Guarner, F. and J.-R. Malagelada, Gut flora in health and disease. The Lancet, 2003. 361(9356): p. 512-519.
61. Wadood, A., et al., Subtractive genome analysis for in silico identification and characterization of novel drug targets in Streptococcus pneumonia strain JJA. Microbial pathogenesis, 2017.
62. Savage, D.C., Microbial ecology of the gastrointestinal tract. Annual Reviews in Microbiology, 1977. 31(1): p. 107-133.
63. Peterson, J., et al., The NIH human microbiome project. Genome research, 2009. 19(12): p. 2317-2323.

64. Methe, B.A., et al., A framework for human microbiome research. Nature, 2012. 486(7402): p. 215.
65. Szklarczyk, D., et al., The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic acids research, 2016: p. gkw937.
66. Kuntal, B.K., P. Aparoy, and P. Reddanna, EasyModeller: A graphical interface to MODELLER. BMC research notes, 2010. 3(1): p. 226.
67. Xu, D. and Y. Zhang, Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophysical journal, 2011. 101(10): p. 2525-2534.
68. Laskowski, R.A., et al., PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography, 1993. 26(2): p. 283-291.
69. Colovos, C. and T.O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions. Protein science, 1993. 2(9): p. 1511-1519.
70. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD Bairoch. A Protein Identification and Analysis Tools on the ExPASy Server. Nucleic Acids Res. 2003;31:3784-3788.
71. Buus S, Lauemoller S, Worning P, Kesmir C, Frimurer T, Corbet S, et al. Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee'artificial neural network approach. HLA. 2003;62:378-84.
72. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Journal of molecular biology. 2001;305:567-80.
73. Dimitrov I, Bangov I, Flower DR, Doytchinova1 I. AllerTOP v.2- a server for in silico prediction of allergens. J. of Molecular Modeling. 2013;20:2278.
74. Dimitrov I, Naneva L, Doytchinova1 I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. Bioinformatics. 2014;30: 846–851.
75. Chrysostomou C and Seker H. Prediction of protein allergenicity based on signal-processing bioinformatics approach. 36th annual international conference of the IEEE engineering in medicine and biology society. 2014;2014:808-11.

76. Fiers M WEJ, Kleter GA, Nijland H, Peijnenburg A, Nap JP and Ham R. Allermatch™, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. BMC Bioinformatics. 2004;5:133.
77. Kolaskar A, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS letters. 1990;276:172-4.
78. Emini EA, Hughes JV, Perlow D, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. Journal of virology. 1985;55:836-9.
79. Karplus P, Schulz G. Prediction of chain flexibility in proteins. Naturwissenschaften. 1985;72:212-3.
80. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic acids research. 2017;45:W24-W9.
81. Chou P, Fasman G. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. 1978;47:45–148.
82. Parker J, Guo D and Hodges R. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry. 1986;25:5425-32.
83. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, Lundegaard C, Sette A, Lund O, Bourne PE, Nielsen M. Immune epitope database analysis resource. Nucleic acids research. 2012 May 18;40(W1):W525-30.
84. Rana A, Akhter, YA. multi-subunit based, thermodynamically stable model vaccine using combined immunoinformatics and protein structure based approach. Immunobiology. 2016;221:544–557.
85. Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatabadi SS and Ghasemi Y. Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against Staphylococcus aureus. Infect. Genet. Evol. 2017;48:83–94.
86. Pandey RK, Sundar S, Prajapati VK. Differential expression of miRNA regulates T cell differentiation and plasticity during visceral leishmaniasis infection. Front. Microbiol. 2016;7:206.

87. Hasan M, Hakim A, Iqbal A, Bhuiyan FR, Begum MK, Sharmin S and Abir RA (2015) Computational study and homology modeling of phenol hydroxylase: key enzyme for phenol degradation. Int J ComputBioinfo In Silico Model 4(4):691-698

88. Kosciolek T, Jones DT. De novo structure prediction of globular proteins aided by sequence variation-derived contacts. PloS One. 2014;9(3):e92197.

89. Petersen B, Lundegaard C, Petersen TN. NetTurnP–neural network prediction of beta-turns by use of evolutionary information and predicted protein sequence features. PLoS One. 2010;5(11):e15079.

90. Thaysen-Andersen M, Packer NH. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. Glycobiology. 2012;22(11):1440-52.

91. Kallberg M, Margaryan G, Wang S, Ma J and Xu J. RaptorX server: a resource for template-based protein structure modeling. InProtein Structure Prediction. 2014;1137:17-27.

92. Peng J, Xu J. RaptorX: exploiting structure information for protein alignment by statistical inference. Proteins: Structure, Function, and Bioinformatics. 2011;79(S10):161-71.

93. Xu D and Zhang Y. Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization. Biophysical Journal. 2011;101:2525-2534.

94. Zhang J, Liang Y, Zhang Y. Atomic-Level Protein Structure Refinement Using Fragment-Guided Molecular Dynamics Conformation Sampling. Structure 2011;19:1784-1795.

95. Lovell SC, Davis IW, Arendall WB, Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi, psi and beta deviation. Proteins: Structure, Function & Genetics. 2002;50:437-450.

96. Hasan M, Joy ZF, Bhuiyan EH, Islam MS (2015) In Silico Characterization and Motif Election of Neurotoxins from Snake Venom. American Journal of Biochemistry & Biotechnology 11(2):84

97. Craig DB, Dombkowski AA. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. BMC Bioinformatics. 2013;14:346.
98. Solanki V, Tiwari V. Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against Acinetobacter baumannii. Scientific reports. 2018;8(1):9044.
99. Newton HJ, Ang DK, van Driel IR, Hartland EL. Molecular pathogenesis of infections caused by Legionella pneumophila. Clinical microbiology reviews. 2010 Apr 1;23(2):274-98.
100. Ahmadishoar S, Kariminik A. Toll-like receptor 2 and its roles in immune responses against Legionella pneumophila. Life sciences. 2017 Nov 1;188:158-62.
101. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein-protein docking. Nucl. Acids Res. 2004; 32: W96–9.
102. Macalino S, Basith S, Clavio N, Chang H, Kang S, Choi S. Evolution of In Silico Strategies for Protein-Protein Interaction Drug Discovery. Molecules 2018; 23(8): 1963.
103. Extra- Kangueane P, Nilofer C. Protein-Protein Docking: Methods and Tools. InProtein-Protein and Domain-Domain Interactions. 2018:161-168.
104. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucl. Acids Res. 2005;33:W363–7.
105. Aalten DMF, Groot BL, Findlay JBC, Berendsen HJC, Amadei A. A Comparison of Techniques for Calculating Protein Essential Dynamics. Journal of Computational Chemistry. 1997;18(2):169-181.
106. Wuthrich K, Wagner G, Rene Richarz, and Werner Braun. Correlations between internal mobility and stability of globular proteins, Biophys. J. 1980;549-558.
107. Tama F, Brooks CL. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. Annu. Rev. Biophys. Biomol. Struct. 2006;35:115-33.
108. Cui Q, Bahar I. normal mode analysis theoretical and applications to biological and chemical systems. Briefing in Bioinformatics. 2007;8(5):378-379.
109. Lopez-Blanco JR, Garzon JI and Chacon P. iMod: multipurpose normal mode analysis in internal coordinates. Bioinformatics. 2011;27(20):2843-2850.
110. Lopez-Blanco JR, Aliaga JI, Quintana-Orti ES, Chacon P. iMODS: internal coordinates normal mode analysis server. Nucleic Acids Res. 2014;42:W271–W276.

111. Awan FM, Obaid A, Ikram A, Janjua HA. Mutation-structure function relationship based integrated strategy reveals the potential impact of deleterious missense mutations in autophagy related proteins on hepatocellular carcinoma (HCC): a comprehensive informatics approach. Int. J. Mol. Sci. 2017;18(1):139.
112. Prabhakar PK, Srivastava A, Rao KK and Balaji PV. Monomerization alters the dynamics of the lid region in campylobacter jejuniCstII: an MD simulation study. J. Biomol. Struct. Dyn. 2016;34(4):778–79.
113. Grote, A. et al. J Cat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic acids research. 2005; 33:W526–531.
114. Hasan M, Islam S, Chakraborty S, Mustafa AH, Azim KF, Joy ZF, Hossain MN, Hossain S and Hasan MN. Contriving a chimeric polyvalent vaccine to prevent infections caused by herpes simplex virus (type-1 and type-2): an exploratory immunoinformatic approach. J BiomolStructDyn. 2019. DOI: 10.1080/07391102.2019.1647286
115. Bartram, J., Chartier, Y., Lee, J. V., Pond, K., &Surman-Lee, S. (Eds.). (2007). *Legionella and the prevention of legionellosis*. World Health Organization.
116. Yoo BK et al. (2004). Gemifloxacin: a new fluoroquinolone approved for treatment of respiratory infections. *Annals of Pharmacotherapy*, 38:1226–1235.
117. Rahimi B, Vesal A (2017). Antimicrobial Resistance Properties of Legionella pneumophila Isolated From The Cases of Lower Respiratory Tract Infections. Biomed. Pharma. J.; **10**(1), 59-65.
118. Hediger, M.A., E. Turk, and E.M. Wright, Homology of the human intestinal Na+/glucose and Escherichia coli Na+/proline cotransporters. Proceedings of the National Academy of Sciences, 1989. 86(15): p. 5748-5752.
119. Swango, K., et al., Amino acid homologies between human biotinidase and bacterial aliphatic amidases: putative identification of the active site of biotinidase. Molecular genetics and metabolism, 2000. 69(2): p. 111-115.
120. Zhang, R., H.Y. Ou, and C.T. Zhang, DEG: a database of essential genes. Nucleic acids research, 2004. 32(suppl_1): p. D271-D272.
121. Judson, N. and J.J. Mekalanos, TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. Nature biotechnology, 2000. 18(7): p. 740.

122. Michael, C. A., Dominey-Howes, D. & Labbate, M. The antimicrobial resistance crisis: causes, consequences, and management. *Front. Public Health***2**, 145 (2014).
123. Baliga, P., Shekar, M. &Venugopal M. N. Potential Outer Membrane Protein Candidates for Vaccine Development Against the Pathogen Vibrio anguillarum: A Reverse Vaccinology Based Identification. *Curr. Microbiol.***75**, 368-77 (2018).
124. Golkar, Z., O. Bagasra, and D.G. Pace, Bacteriophage therapy: a potential solution for the antibiotic resistance crisis. The Journal of Infection in Developing Countries, 2014. 8(02): p. 129-136.
125. Bartlett, J.G., D.N. Gilbert, and B. Spellberg, Seven ways to preserve the miracle of antibiotics. Clinical Infectious Diseases, 2013. 56(10): p. 1445-1450.
126. Michael, C.A., D. Dominey-Howes, and M. Labbate, The antimicrobial resistance crisis: causes, consequences, and management. Frontiers in public health, 2014. 2: p. 145.
127. Fung, M., et al., Evaluation of the Characteristics of Safety Withdrawal of Prescription Drugs from Worldwide Pharmaceutical Markets-1960 to 1999. Drug Information Journal, 2001. 35(1): p. 293-317.
128. Moses, P.L., et al., Severe hepatotoxicity associated with bromfenac sodium. The American journal of gastroenterology, 1999. 94(5): p. 1393.
129. Andrade, R.J., et al., Acute liver injury associated with the use of ebrotidine, a new H2-receptor antagonist. Journal of hepatology, 1999. 31(4): p. 641-646.
130. Lazarczyk D, Goldstein N, Gordon S. (2000) Trovafloxacin hepatotoxicity. The American journal of gastroenterology 95(9):2585.
131. Ghaffari-Nazari, H. et al. (2015). Improving Multi-Epitope Long Peptide Vaccine Potency by Using a Strategy that Enhances CD4+T Helpin BALB/c Mice. PloS one 10:e0142563.
132. In silico epitope prediction and immunogenic analysis for penton base epitope-focused vaccine against hydropericardium syndrome in chicken.
133. Yang, Y. et al. (2015) In silico design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. Human Vaccines &Immunotherapeutics 11:795–805.
134. Khan M, Khan S, Ali A, Akbar H, Sayaf AM, Khan A, Wei DQ. Immunoinformatics approaches to explore Helicobacter Pylori proteome (Virulence Factors) to design B and T cell multi-epitope subunit vaccine. Scientific reports. 2019 Sep 16;9(1):1-3.

Table Captions

Table 1: Subtractive genomic analysis scheme towards the identification of novel therapeutic targets

Table 2: Proteins involved in pathogen specific pathways

Table 3: Identified druggable targets with drug names

Table 4: List of probable antigenic proteins for vaccine targets

Table 5: Human microflora non-homology analysis and subcellular localization

Table 6: List of identified novel drug targets against *Legionella pneumophila*

Table7: Predicted T-cell epitopes (MHC-I peptides and MHC-II peptides) of LEGPH protein TolB

Table 8: Allergenicity pattern of the predicted B-cell epitopes generated from LEGPH protein TolB

Table 9: Allergenicity prediction and antigenicity analysis of designed vaccine constructs

Table 10: Docking score of vaccine construct V1 with different HLA alleles including HLA-DRB1*03:01 (1A6A), (HLA-DRB5*01:01 (1H15), HLA-DRB3*01:01 (2Q6W), HLA-DRB1*04:01 (2SEB), HLA-DRB1*01:01 (2FSE), and HLA-DRB3*02:02 (3C5J)

Supplementary Tables

Supplementary Table 1: KO assigned (KEGG orthlogy) proteins not involved in any metabolic pathway

Supplementary Table 2: Functional family classification of the hypothetical proteins by SVMProt server

Supplementary Table 3: Templates used for homology modeling by EasyModeller 4.0

Supplementary Table 4: ProtParam analysis of LEGPH protein TolB

Supplementary Table 5: Identified epitope clusters among the top CTL, HTL and BCL epitopes

Figure captions

Fig. 1. Proteome exploration of *Legionella pneumophila* to identify novel drug targets.

Fig. 2. Flow chart summarizing the protocols for the prediction of epitope based vaccine candidate by *in silico* reverse vaccinology technique.

Fig. 3. Investigation of Protein-Protein Interactions throughSTRING v10.5 server, **A:**UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase (murE), **B:**Trigger factor (tig), **C:** Lipid II flippaseMurJ (mviN), **D:** Potassium transport system protein kup 1 (kup1), **E:**Protein TolB (tolB)

Fig. 4. Population coverage analysis of LEGPH protein TolB

Fig. 5. Tertiary structure prediction and validation of vaccine protein V1, **A:** Cartoon format **B:** Validation of the 3D structure of vaccine protein V1 by Ramachandran plot analysis.

Fig. 6. Disulfide engineering of vaccine protein V1; **A:** Initial form, **B:** Mutant form.

Fig. 7. Docked complex of vaccine construct V1 with human TLR8; **A:** Cartoon format and **B:** Ball structure.

Fig. 8. Molecular dynamics simulation of vaccine protein V1-TLR2 complex; stability of the protein-protein complex was investigated through mobility **(A)**, eigenvalue **(B)**, B-factor **(C)**, deformability **(D)**, covariance **(E)** and elastic network **(F)** analysis.

Fig. 9. Restriction digestion **(A)** and *in silico* cloning **(B)** of the gene sequence of final vaccine construct V1 into pET28a(+) expression vector. Target sequence was inserted between HindIII (173) and BamHI (198).

Supplementary Figures

Supplementary Fig. 1. Tertiary structure prediction of the vaccine target through EasyModeller 4.0 **(A)** and Ramachandran Plot analysis **(B)**.

Supplementary Fig. 2. Quality factor analysis of modeled 3D structure of LEGPH protein TolB by ERRAT

Supplementary Fig. 3. Prediction of B cell linear epitope and intrinsic properties for LEGPH protein Tol B using different scales (**A:**Bepipred, **B:** Surface accessibility, **C:**Emini surface, **D:** Flexibility, **E:** Antigenicity, **F:** Hydrophilicity). **(**For each graph: x-axis and y-axis represent the position and score; residues that fall above the threshold value are shown in yellow color; the highest peak in yellow color identifies most favored position).

Supplementary Fig. 4. Secondary structure prediction of designed vaccine V1 using PESIPRED server

Supplementary Fig. 5. 3D modelled structure of vaccine protein V2 **(A)** and V3 **(B)** generated via RaptorX server.

Supplementary Files

Supplementary File 1: A total 210 'anti-targets' of human

Supplementary File 2: Unique pathways of Legionella pneumophila retrieved from KEGG server.

Supplementary File 3: Subcellular localization of nonhomologous essential proteins involved in unique pathways of Legionella pneumophila

Supplementary File 4: Details of Protein-Protein Interactions through STRING v10.

Table 1: Subtractive genomic analysis scheme towards the identification of novel therapeutic targets

No.	***Steps***	***Legionella pneumophila subsp. pneumophila (strain Philadelphia / ATCC 33152 / DSM 7513)***
1	*Total number of proteins*	2930
2	Removed paralogous (>60% identical) in CD-Hit	2878
3	Number of proteins nonhomologous to *H. sapiens* using BLASTp (E value 10^{-3})	2010
4	Essential proteins in DEG 15.2 server (E value $\leq 10^{-100}$, Bit score >100)	125
5	Essential Proteins involved only in unique metabolic pathways (KAAS at KEGG)	4
6	Proteins assigned KO (KEGG Orthology) but not assigned in any pathway	29
7	Hypothetical essential proteins assigned KO (KEGG Orthology) but not assigned in any pathway	3
8	Essential membrane proteins using PSORTb, CELLO, ngLOC, PSLpred servers	12
9	Essential cytoplasmic proteins using PSORTb, CELLO, ngLOC, PSLpred servers	13
	Proteins of unknown location	1
10	Essential proteins found to be novel in DrugBank 5.1.0 (using default parameters)	25

11	Proteins non-homologous to ‘anti-targets’ using BLASTp (E value <0.005, Identity <30%, Query length >30%)	25
12	Membrane proteins showing antigenicity using VexiJen v2.0 (Threshold value >0.4)	12
13	Number of cytoplasmic and membrane proteins less similar to Human microbiota using BLASTp against "43021[BioProject]" (E value 1, Identity<47%)	5
14	Proteins selected for epitope design	1
15	**Novel drug targets**	**3**

Table 2: Proteins involved in pathogen specific pathways

S. no.	***KO assignment***	***Accession no.***	***Protein name***	***Pathway name***
1	*K18138*	tr|Q5ZXL1	Multidrug resistance protein	beta-Lactam resistance; Cationic antimicrobial peptide (CAMP) resistance
2	K01928	tr|Q5ZX16	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase	Lysine biosynthesis; Peptidoglycan biosynthesis
3	*K03587*	tr|Q5ZX17	Peptidoglycan D,D-transpeptidaseFtsI	Peptidoglycan biosynthesis; beta-Lactam resistance
4	*K05515*	tr|Q5ZVR5	Peptidoglycan D,D-transpeptidaseMrdA	Peptidoglycan biosynthesis; beta-Lactam resistance

Table 3: Identified druggable targets with drug names

Sl. No.	***Accession No.***	***Protein name***	***Drugbank Id***	***Drug name***	***Bit score***	***Query length***
1	tr|Q5ZS52	RNA polymerase sigma factor RpoH	DB08874	Fidaxomicin	103.605	318
2	tr|Q5ZT04	RNA polymerase sigma factor RpoD	DB08226, DB08266	Myxopyronin B, Methyl [(1E,5R)-5-{3-[(2E,4E)-2,5-dimethyl-2,4-octadienoyl]-2,4-dioxo-3,4-dihydro-2H-pyran-6-yl}hexylidene]carbamate	302.368	248
3	tr|Q5ZUP0	UDP-N-acetylmuramate--L-alanyl-gamma-D-glutamyl-meso-2,6-diaminoheptandioate ligase	DB01673, DB03909, DB04395	Uridine-5'-Diphosphate-N-Acetylmuramoyl-L-Alanine, Adenosine-5'-[Beta, Gamma-Methylene]Triphosphate, Phosphoaminophosphonic Acid-Adenylate Ester	145.591	398
4	tr|Q5ZXL1	Multidrug resistance protein	DB03825, DB04209, DB03619	Rhodamine 6G, Dequalinium, Deoxycholic Acid	427.557	1020
5	tr|Q5ZX17	Peptidoglycan D,D-transpeptidaseFtsI	DB00303, DB00671	Ertapenem, Cefixime	122.094	632

6	tr|Q5ZVR5	Peptidoglycan D,D-transpeptidaseMrdA	DB01598, DB01329, DB01327, DB01163, DB01328, DB01413, DB01415, DB00948, DB00438, DB00303, DB06211	Imipenem, Cefoperazone, Cefazolin, Amdinocillin, Cefonicid, Cefepime, Ceftibuten, Mezlocillin, Ceftazidime, Ertapenem, Doripenem	486.493	612
7	tr|Q5ZZI4	Amino acid permease	DB00123, DB00125, DB00129	L-Lysine, L-Arginine, Ornithine	155.606	**424**

Table 4: List of probable antigenic proteins for vaccine targets

S. no.	***Accession no.***	***Vaxijen v2.0 score***	***Subcellular localization***
1	tr|Q5ZS84	0.4733	Inner Membrane
2	tr|Q5ZUM7	0.4869	Inner Membrane
3	tr|Q5ZT44	0.5906	Inner Membrane
4	tr|Q5ZVR6	0.6058	Inner Membrane
5	sp|Q5ZV69	0.6714	Outer membrane
6	sp|Q5ZW98	0.6393	Inner Membrane
7	sp|Q5ZSY2	0.4632	Inner Membrane
8	sp|Q5ZTN5	0.5461	Inner Membrane
9	sp|Q5ZRL2	0.4190	Inner Membrane
10	tr|Q5ZY66	0.4935	Outer membrane
11	tr|Q5ZX71	0.5807	Inner Membrane

Table 5: Human microflora non-homology analysis and subcellular localization

S. no.	Accession no.	Microbiome similarity	Subcellular localization
1	tr|Q5ZX16	<46	Cytoplasmic
2	sp|Q5ZUD8	<45	Cytoplasmic
3	tr|Q5ZS84	<47	Inner Membrane
4	sp|Q5ZV69	<47	Outer membrane
5	**sp|Q5ZW98**	**<44**	**Inner Membrane**

Table 6: List of identified novel drug targets against *Legionella pneumophila*

S. no.	***Accession no.***	***Protein names***
1	tr|Q5ZX16	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase
2	sp|Q5ZUD8	Trigger factor
3	tr|Q5ZS84	Probable lipid II flippaseMurJ

Table 7: Predicted T-cell epitopes (MHC-I peptides and MHC-II peptides) of LEGPH protein TolB

Epitope		***Length***	***Antigenicity score***	***Conservancy***	***Toxicity***	***Allergenicity***
MHC-I peptides (CTL epitopes)	IISLFLLLF	9	2.8937	88.89%	Non-Toxin (-0.82)	Non-allergen
	TSGRGGSPQ	9	2.5779	100%	Non-Toxin (-0.86)	Non-allergen
	TSGRGGSPQV	10	2.5666	100%	Non-Toxin (-0.90)	Non-allergen
	SGRGGSPQV	9	2.4721	100%	Non-Toxin (-0.83)	Non-allergen
	IGVQNTGGG	9	2.2694	83.33%	Non-Toxin (-0.77)	Non-allergen
	RIISLFLLLF	10	2.2479	88.89%	Non-Toxin (-0.91)	Non-allergen
	VIALDLELT	9	2.0381	100%	Non-Toxin (-1.30)	Non-allergen
	ISLFLLLFT	9	2.0361	88.89%	Non-Toxin (-0.99)	Non-allergen
	SGRGGSPQVY	10	2.0168	100%	Non-Toxin (-1.04)	Non-allergen
	IISLFLLLFT	10	2.0107	88.89%	Non-Toxin (-0.82)	Non-allergen
MHC-IIpeptides (HTL epitopes)	TSGRGGSPQVYRLSL	15	1.8097	100%	Non-Toxin (-1.42)	Non-allergen
	GGRSRYSLEVADADG	15	1.6123	100%	Non-Toxin (-0.57)	Non-allergen
	SGRGGSPQVYRLSLA	15	1.5498	100%	Non-Toxin (-1.46)	Non-allergen
	TFGNSIDTEPRYSPD	15	1.4660	100%	Non-Toxin (-1.13)	Non-allergen
	AREGDVQEPAWSPYL	15	1.4197	88.89%	Non-Toxin (-0.76)	Allergen
	YSLEVADADGHNPQS	15	1.3878	100%	Non-Toxin (-0.65)	Non-allergen
	QLTFGNSIDTEPRYS	15	1.3645	100%	Non-Toxin (-1.45)	Allergen
	PKIYDVDLSSGSMKQ	15	1.3508	72.22%	Non-Toxin (-0.67)	Non-allergen
	LTFGNSIDTEPRYSP	15	1.3006	100%	Non-Toxin (-1.19)	Non-allergen
	YDVDLSSGSMKQLTF	15	1.2975	72.22%	Non-Toxin (-0.77)	Non-allergen

Table 8: Allergenicity pattern of the predicted B-cell epitopes generated from LEGPH protein TolB

Protein	method	Start	End	Peptide	Length	Allergenicity
LEGPH Protein TolB	Bepipred linear epitope prediction	66	77	SGPQGANSQSSV	12	Non Allergen
		404	417	AREGDVQEPAWSPY	14	Allergen
	Chou &Fasman beta turn prediction	194	200	SWSPNGK	7	Non Allergen
		294	300	SGRGGSP	7	Non Allergen
	Emini surface accessibility prediction	206	212	SFEKKKA	7	Non Allergen
		277	287	IDTEPRYSPDG	11	Non Allergen
	Karplus& Schulz flexibility prediction	160	166	QRNGGRS	7	Allergen
		308	314	ADGQISR	7	Non Allergen
	Kolaskar&Tongaonkar antigenicity scale	5	23	IISLFLLLFTGQVIALDLE	19	Non Allergen
		243	252	GQHLAVVLSK	10	Non Allergen
	Parker hydrophilicity prediction	174	180	DADGHNP	7	Allergen
		283	289	YSPDGRS	7	Non Allergen

Table 9: Allergenicity prediction and antigenicity analysis of designed vaccine constructs

Vaccine Constructs	Composition	Complete Sequence of Vaccine Constructs	Allergenicity	VaxiJen score (Threshold 0.4)
V1	Predicted CTL, HTL & BCL epitopes of LEGPH Protein TolBwith β defensin adjuvant &PADRE sequence	**EAAAK**GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKK**EAAAK**AKFVAAWTLKAAA**GGGS**RIISLFLLLFT**GGGS**TSGRGGSPQVY**GGGS**QVIALDLELTQ**GGGS**IGVQNTGGGP**GGGS**LPAREGDVQ**GGGS**DVDLSSGSMK**GGGS**TFGNSIDTE**GPGPG**TPKIYDVDLSSGSMKQLTFG**GPGPG**LTFGNSIDTEPRYSPDG**GPGPG**FTSGRGGSPQVYRLSLA**GPGPG**YSLEVADADGHNPQS**GPGPG**RVTFEGNYNARASYT**GPGPG**GGRSRYSLEVADADG**GPGPG**ISGPQGANSQSSVST**GPGPG**VIALDLELTQGINSA**KK**SGPQGANSQSSV**KK**AREGDVQEPAWSPY**KK**SWSPNGK**KK**SGRGGSP**KK**SFEKKKA**KK**IDTEPRYSPDG**KK**ADGQISR**KK**IISLFLLLFTGQVIALDLE**KK**GQHLAVVLSK**KK**YSPDGRS**KK**AKFVAAWTLKAAA**GGGS**	Non Allergen	0.6034
V2	Predicted CTL, HTL & BCL epitopes of LEGPH Protein TolBwith	**EAAAK**MAKLSTDELLDAFKEMTLLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILEAAGDKKIGVIKVVREIVSGLGLKEAKDLVDGAPKPLLEKVAKEAADEAKAKLEAAGATVTVK**EAAAK**AKFVAAWTLKAAA**GGGS**RIISLFLLLFT**GGGS**TSGRGGSPQVY**GGGS**QVIALDLELTQ**GGGS**IGV	Non Allergen	0.5524

	L7/L12 ribosomal protein adjuvant & PADRE sequence	QNTGGGP**GGGS**LPAREGDVQ**GGGS**DVDL SSGSMK**GGGS**TFGNSIDTE**GPGPG**TPKIYD VDLSSGSMKQLTFG**GPGPG**LTFGNSIDTEP RYSPDG**GPGPG**FTSGRGGSPQVYRLSLA**G PGPG**YSLEVADADGHNPQS**GPGPG**RVTFE GNYNARASYT**GPGPG**GGRSRYSLEVADA DG**GPGPG**ISGPQGANSQSSVST**GPGPG**VIA LDLELTQGINSA**KK**SGPQGANSQSSV**KK**A REGDVQEPAWSPY**KK**SWSPNGK**KK**SGRG GSP**KK**SFEKKKA**KK**IDTEPRYSPDG**KK**AD GQISR**KK**IISLFLLLFTGQVIALDLE**KK**GQH LAVVLSK**KK**YSPDGRS**KK**AKFVAAWTLK AAA**GGGS**		
V3	Predicted CTL, HTL & BCL epitopes of LEGPH Protein TolBwith HABA adjuvant & PADRE sequence	**EAAAK**MAENPNIDDLPAPLLAALGAADLA LATVNDLIANLRERAEETRAETRTRVEERR ARLTKFQEDLPEQFIELRDKFTTEELRKAAE GYLEAATNRYNELVERGEAALQRLRSQTA FEDASARAEGYVDQAVELTQEALGTVASQ TRAVGERAAKLVGIEL**EAAAK**AKFVAAW TLKAAA**GGGS**RIISLFLLLFT**GGGS**TSGRG GSPQVY**GGGS**QVIALDLELTQ**GGGS**IGVQ NTGGGP**GGGS**LPAREGDVQ**GGGS**DVDLS SGSMK**GGGS**TFGNSIDTE**GPGPG**TPKIYDV DLSSGSMKQLTFG**GPGPG**LTFGNSIDTEPR YSPDG**GPGPG**FTSGRGGSPQVYRLSLA**GP GPG**YSLEVADADGHNPQS**GPGPG**RVTFEG NYNARASYT**GPGPG**GGRSRYSLEVADAD G**GPGPG**ISGPQGANSQSSVST**GPGPG**VIAL DLELTQGINSA**KK**SGPQGANSQSSV**KK**AR EGDVQEPAWSPY**KK**SWSPNGK**KK**SGRGG	Non Allergen	0.5726

		SP**KK**SFEKKKA**KK**IDTEPRYSPDG**KK**ADG QISR**KK**IISLFLLLFTGQVIALDLE**KK**GQHL AVVLSK**KK**YSPDGRS**KK**AKFVAAWTLKA AA**GGGS**		

Table 10: Docking score of vaccine construct V1 with different HLA alleles including HLA-DRB1*03:01 (1A6A), (HLA-DRB5*01:01 (1H15), HLA-DRB3*01:01 (2Q6W), HLA-DRB1*04:01 (2SEB), HLA-DRB1*01:01 (2FSE), and HLA-DRB3*02:02 (3C5J)

Vaccine Construct	PDB ID of the HLA Alleles	Global Energy	Hydrogen Bond Energy	ACE	Score	Area
V1	1A6A	-12.98	-4.93	7.18	17648	3028.30
	1H15	-30.47	-5.55	5.82	16608	2414.80
	2Q6W	-10.99	-2.58	7.65	17294	2662.20
	2SEB	-27.99	-8.91	0.83	18036	2334.50
	2FSE	-1.43	-0.46	0.05	19568	2748.70
	3C5J	-9.43	-3.63	15.90	16902	2337.90

Supplementary Table 1:KO assigned (KEGG orthlogy) proteins not involved in any metabolic pathway

S. no.	***KO assignment***	***Accession no.***	***Protein name***
1	K03797	tr|Q5ZY73	Carboxy-terminal protease
2	K03294	tr|Q5ZZI4	Amino acid permease
3	K09822	sp|Q5ZYZ0	Hypothetical protein
4	K07568	sp|Q5ZU04	S-adenosylmethionine:tRNAribosyltransferase-isomerase
5	K06941	sp|Q5ZV93	Dual-specificity RNA methyltransferaseRlmN
6	K03980	tr|Q5ZS84	Probable lipid II flippaseMurJ
7	K03466	tr|Q5ZUM7	Cell division protein FtsK
8	K03089	tr|Q5ZS52	RNA polymerase sigma factor RpoH
9	K07007	tr|Q5ZS92	Hypothetical protein
10	K03656	tr|Q5ZRR4	ATP-dependent DNA helicase Rep
11	K03837	tr|Q5ZT44	Serine transporter
12	K03086	tr|Q5ZT04	LEGPH RNA polymerase sigma factor RpoD
13	K03609	tr|Q5ZUR8	Site-determining protein
14	K02600	tr|Q5ZRV3	Transcription termination/antitermination protein NusA
15	K03722	tr|Q5ZU85	ATP dependent DNA helicase
16	K03592	tr|Q5ZUM2	Peptide maturation protein PmbA
17	K19810	tr|Q5ZYS3	L-lysine 2,3-aminomutase, radical SAM domain protein
18	K05837	tr|Q5ZVR6	PeptidoglycanglycosyltransferaseMrdB

19	K03641	sp|Q5ZV69	Protein TolB
20	K03545	sp|Q5ZUD8	Trigger factor
21	K03549	sp|Q5ZW98	Probable potassium transport system protein kup 1
22	K03549	sp|Q5ZSY2	Probable potassium transport system protein kup
23	K03549	sp|Q5ZTN5	LEGPH Probable potassium transport system protein kup 2
24	K13292	sp|Q5ZRL2	Prolipoproteindiacylglyceryltransferase
25	K06915	tr|Q5ZYU2	Hypothetical protein
26	K07277	tr|Q5ZY66	Outer membrane protein assembly factor BamA
27	K09014	tr|Q5ZXX9	ABC transporter, permease
28	K02198	tr|Q5ZX71	Cytochrome c-type biogenesis protein CcmF
29	K02558	tr|Q5ZUP0	UDP-N-acetylmuramate-L-alanyl-gamma-D-glutamyl-meso-2,6-diaminoheptandioate ligase

Supplementary Table 2: Functional family classification of the hypothetical proteins by SVMProt server

S. no.	***Accession no.***	***KO assignment***	***Family name***	***Probability (%)***	***GO category***
1	sp|Q5ZYZ0	K09822	Iron-binding	98.4	GO:0005506
2	tr|Q5ZS92	K07007	Iron-binding	98.9	GO:0005506
3	**tr|Q5ZYU2**	**K06915**	**All DNA-binding**	**85.4**	**GO:0003677**

Supplementary Table 3: Templates used for homology modeling by EasyModeller 4.0

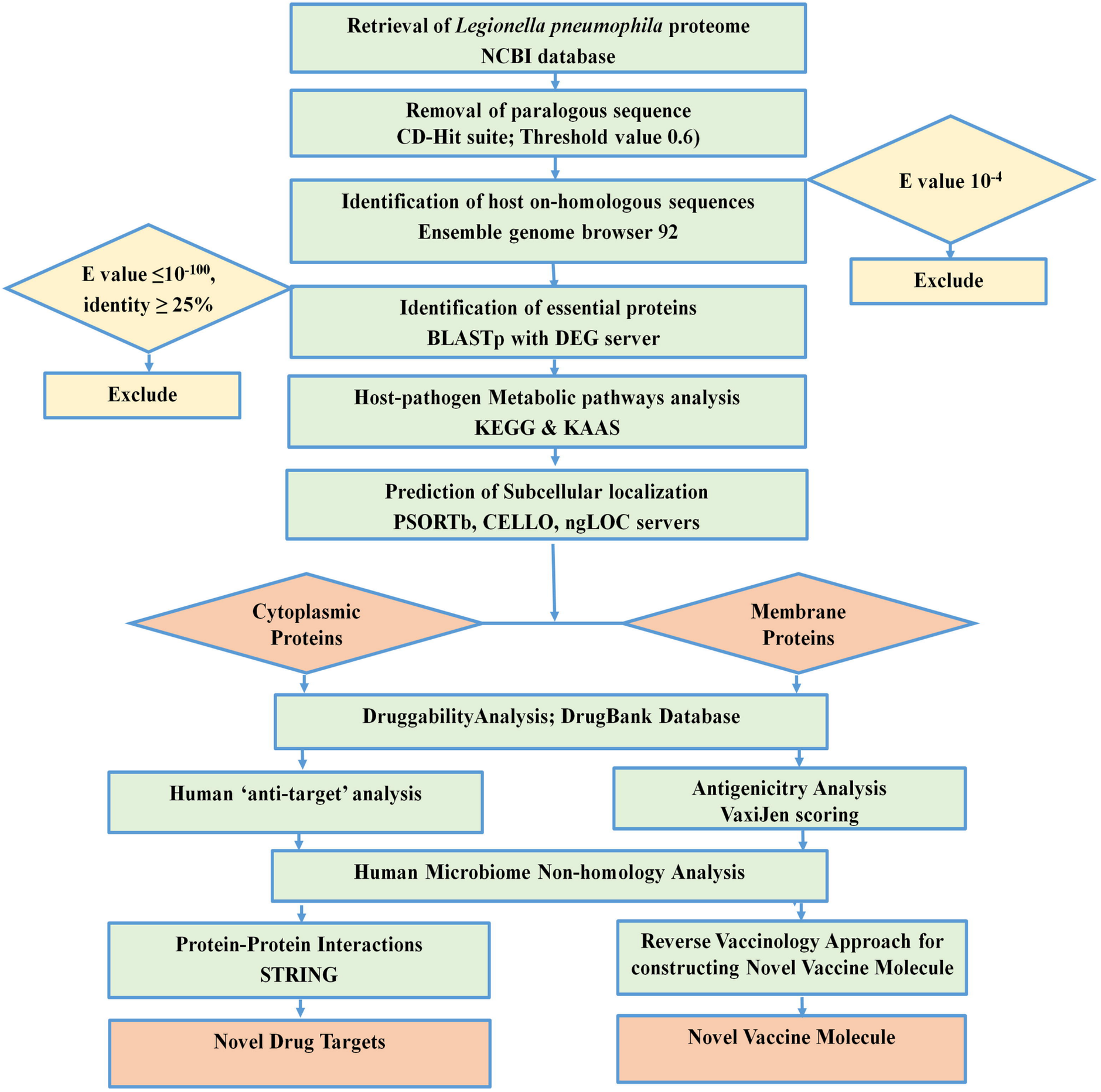
Protein accession no.	PDB ID	Identity (%)	Query cover
	4R40	41	95
sp|Q5ZV69	3IAX	40	99
	1C5K	40	99

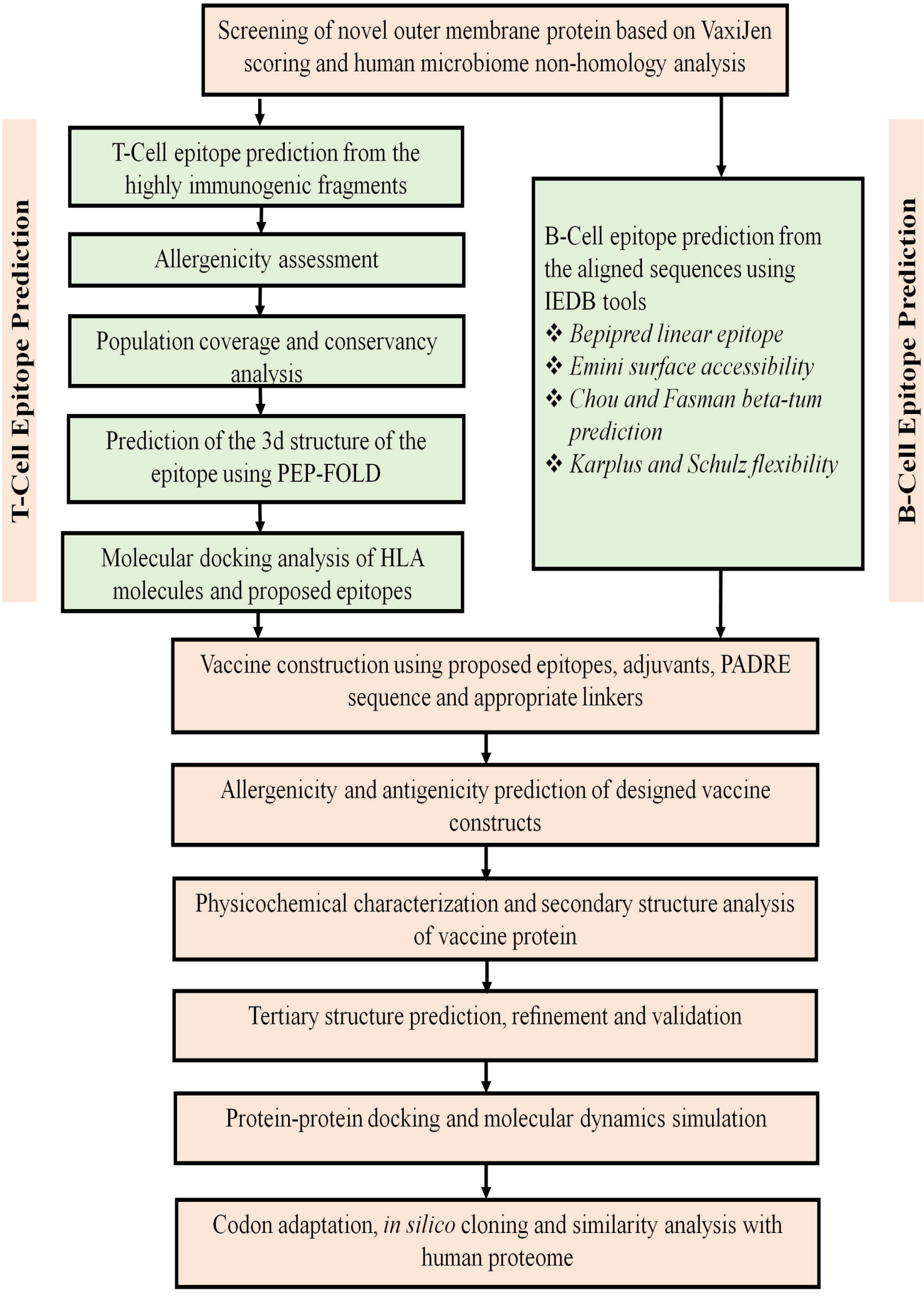
Supplementary Table 4: ProtParam analysis of LEGPH protein TolB

Viral Proteins	***Accession ID***	***Molecular Weight***	***Instability Index***	***Aliphatic Index***	***Theoretical pI***	***No. of Amino acids***	***Estinction Co-Efficient***	***Gravy***
TolB_LEGPH	Q5ZV69	45387.06	38.09	89.81	8.67	419	34380	-0.204

Supplementary Table 5: Identified epitope clusters among top CTL, HTL and BCL epitopes

CTL epitope clusters	**HTL epitope clusters**	**BCL epitope clusters**
RIISLFLLLFT	TPKIYDVDLSSGSMKQLTFG	SGPQGANSQSSV
TSGRGGSPQVY	LTFGNSIDTEPRYSPDG	AREGDVQEPAWSPY
QVIALDLELTQ	FTSGRGGSPQVYRLSLA	SWSPNGK
IGVQNTGGGP	YSLEVADADGHNPQS	SGRGGSP
LPAREGDVQ	RVTFEGNYNARASYT	SFEKKKA
DVDLSSGSMK	GGRSRYSLEVADADG	IDTEPRYSPDG
TFGNSIDTE	ISGPQGANSQSSVST	ADGQISR
	VIALDLELTQGINSA	IISLFLLLFTGQVIALDLE
		GQHLAVVLSK
		YSPDGRS


Retrieval of *Legionella pneumophila* proteome
NCBI database
Removal of paralogous sequence
CD-Hit suite; Threshold value 0.6)
Identification of host on-homologous sequences
Ensemble genome browser 92
E value 10^{-4}
Exclude
E value $\leq 10^{-100}$, identity $\geq 25\%$
Exclude
Identification of essential proteins
BLASTp with DEG server
Host-pathogen Metabolic pathways analysis
KEGG & KAAS
Prediction of Subcellular localization
PSORTb, CELLO, ngLOC servers
Cytoplasmic Proteins
Membrane Proteins
DruggabilityAnalysis; DrugBank Database
Human 'anti-target' analysis
Antigenicitry Analysis
VaxiJen scoring
Human Microbiome Non-homology Analysis
Protein-Protein Interactions
STRING
Reverse Vaccinology Approach for constructing Novel Vaccine Molecule
Novel Drug Targets
Novel Vaccine Molecule

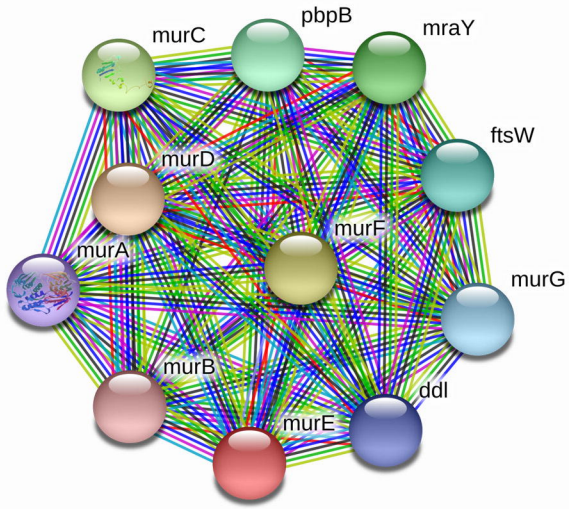


Vaccine Targets

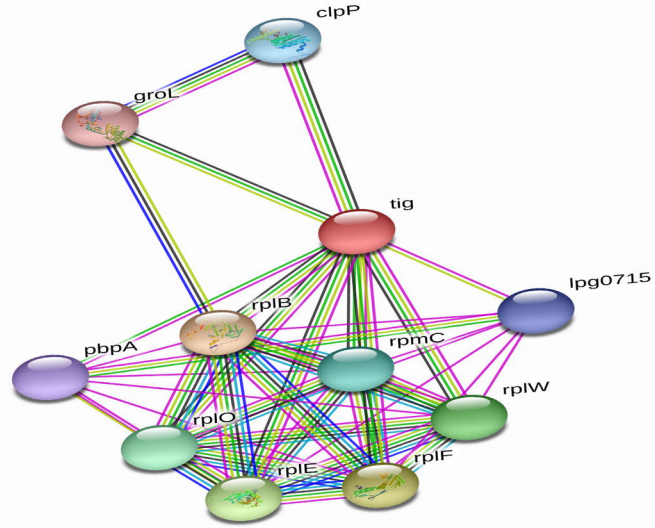
- ❖ Sensor Histidine Protein Kinase (Q87HJ8)
- ❖ Flagellar Hook-associated Protein (Q87JH9)

Resource Used

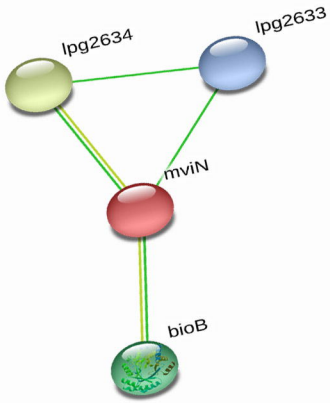
- NCBI, ExPASy
- IEDB
- VaxiJen, TMHMM
- AllerTOP, AllergenFP
- Allermatch, AllergenOnline
- ToxinPred
- MGLTools (AutoDock)
- PyMOL
- PEP-FOLD
- AlgPred
- ProtParam, PSIPRED
- RaptorX, I-TASSER
- ModRefine
- MolProbity
- DbD2
- ClusPro, PatchDock
- iMODS
- JCAT, SnapGene


murC
pbpB
mraY
ftsW
murD
murF
murA
murG
murB
ddl
murE

A


clpP
groL
tig
lpg0715
rplB
pbpA
rpmC
rplW
rplO
rplF
rplE

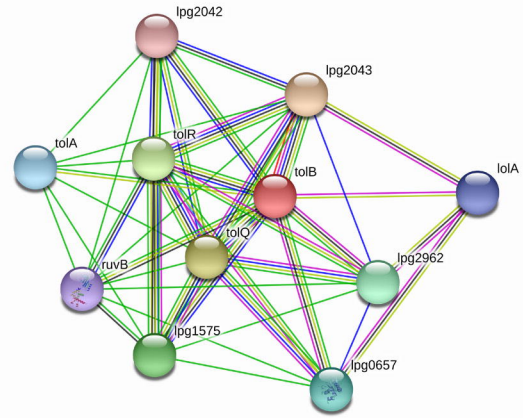
B


lpg2634
lpg2633
mviN
bioB

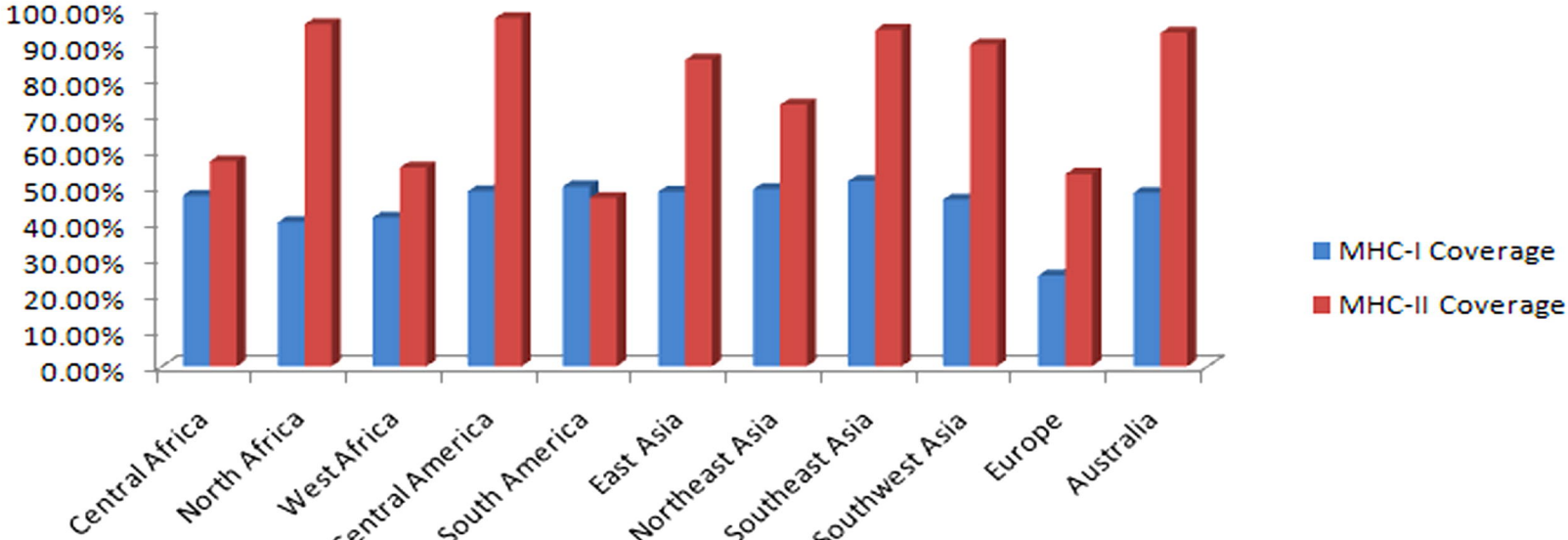
C

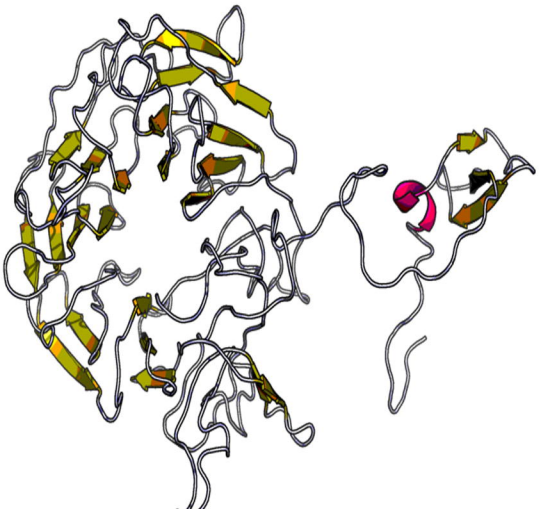
lpg1189
kup1

D

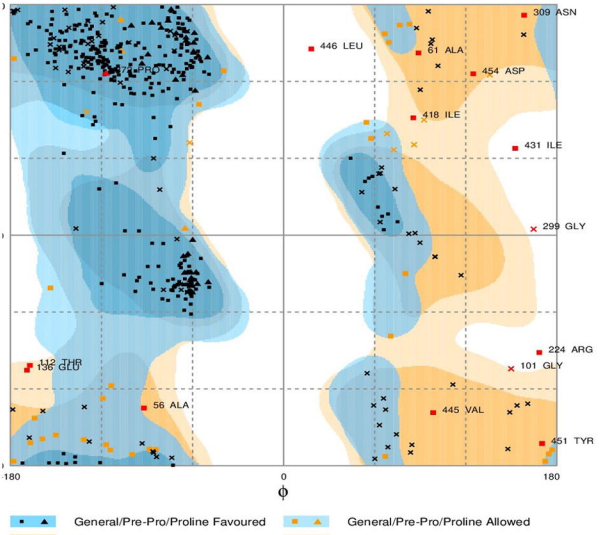

lpg2042
lpg2043
tolA
tolR
tolB
lolA
tolQ
ruvB
lpg2962
lpg1575
lpg0657

E

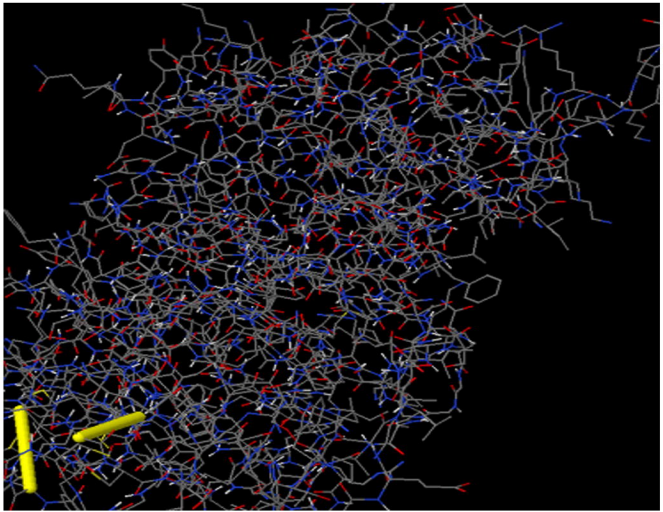

100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
Central Africa
North Africa
West Africa
Central America
South America
East Asia
Northeast Asia
Southeast Asia
Southwest Asia
Europe
Australia
MHC-I Coverage
MHC-II Coverage



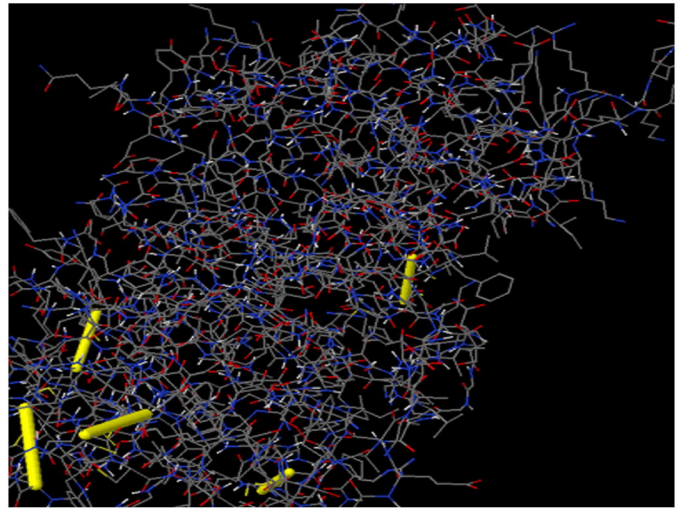
A


309 ASN
446 LEU
61 ALA
454 ASP
418 ILE
431 ILE
299 GLY
224 ARG
101 GLY
112 THR
136 GLU
56 ALA
445 VAL
451 TYR
-180
0
180
ϕ
General/Pre-Pro/Proline Favoured
General/Pre-Pro/Proline Allowed
Glycine Favoured
Glycine Allowed

B

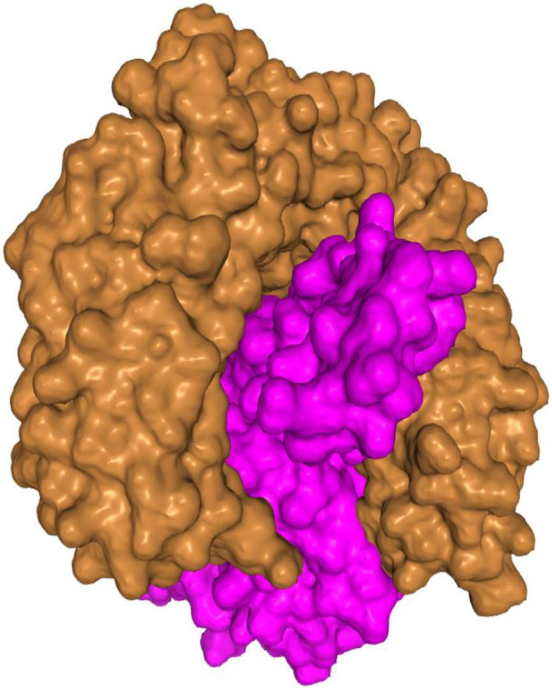


A

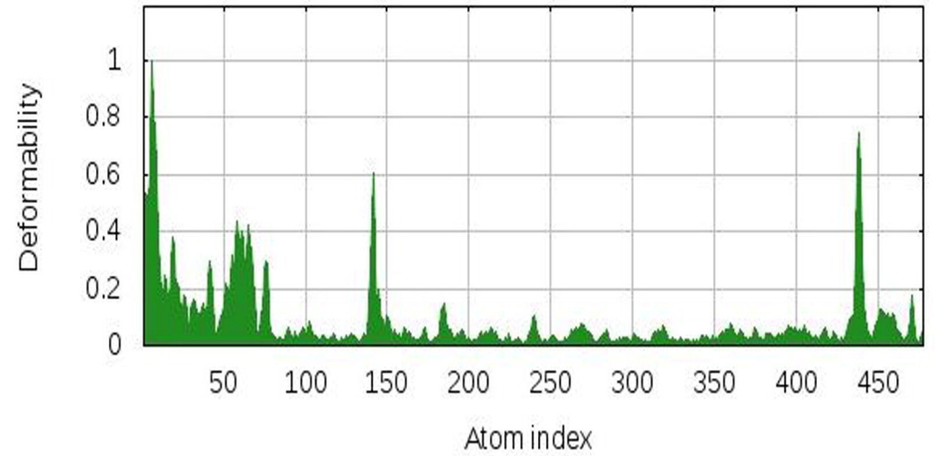


B

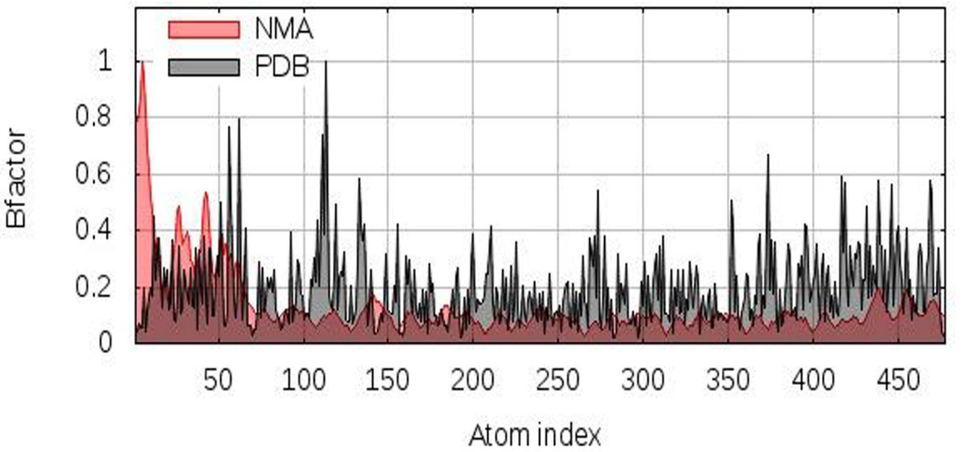
A



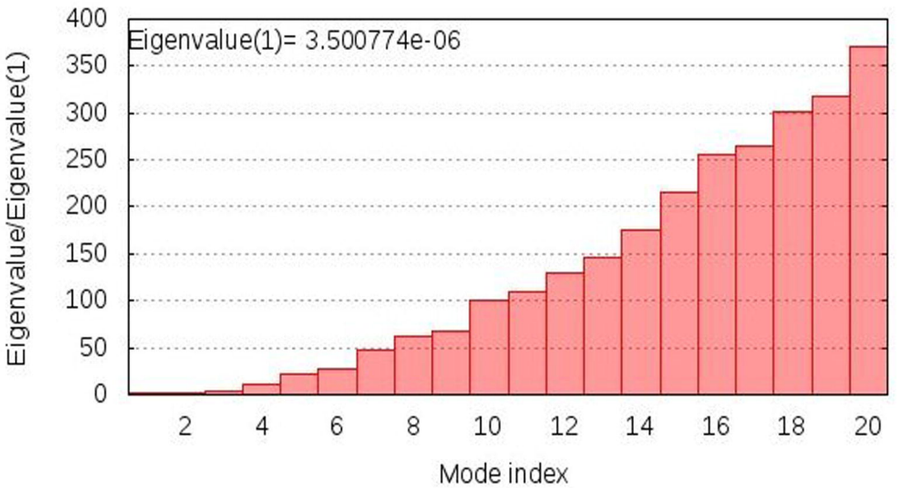
B


Deformability
Atom index

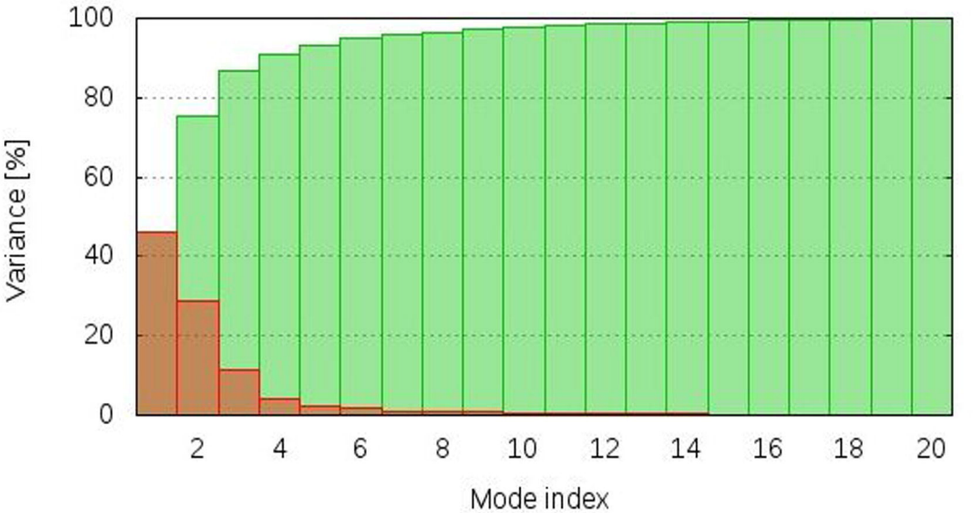
A


NMA
PDB
Bfactor
Atom index

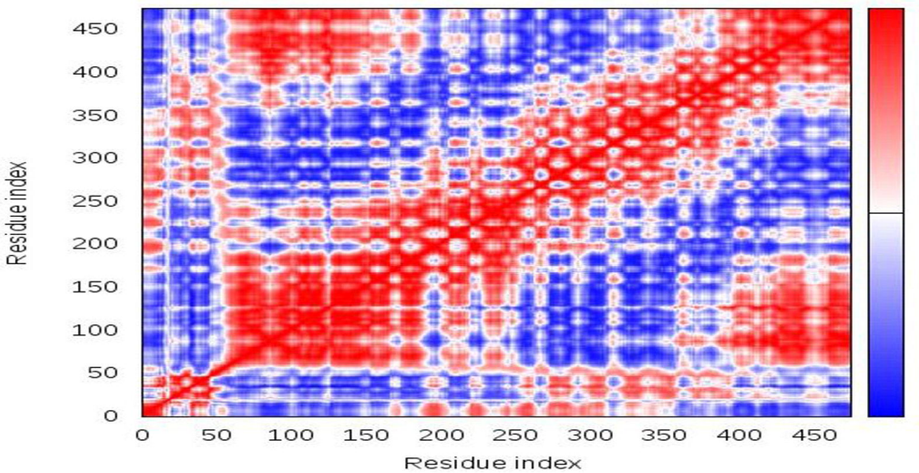
B


Eigenvalue(1)= 3.500774e-06
Eigenvalue/Eigenvalue(1)
Mode index

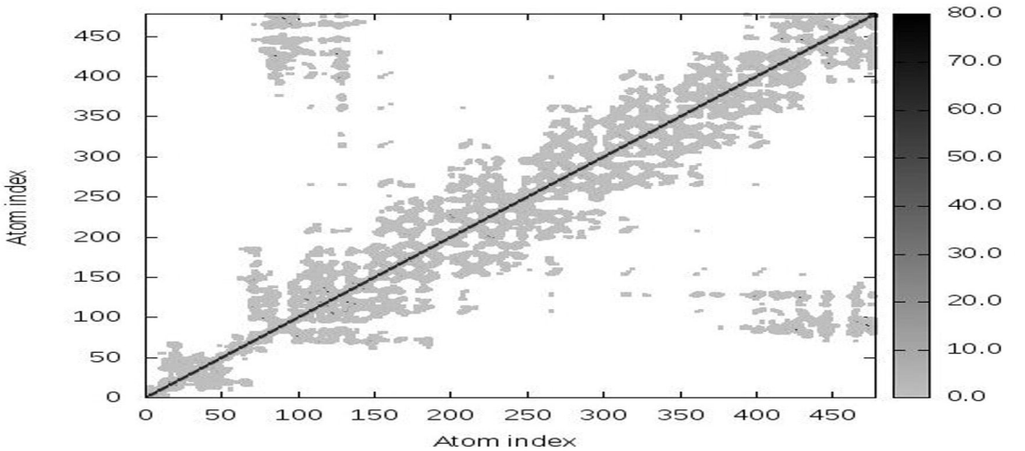
C


Variance [%]
Mode index

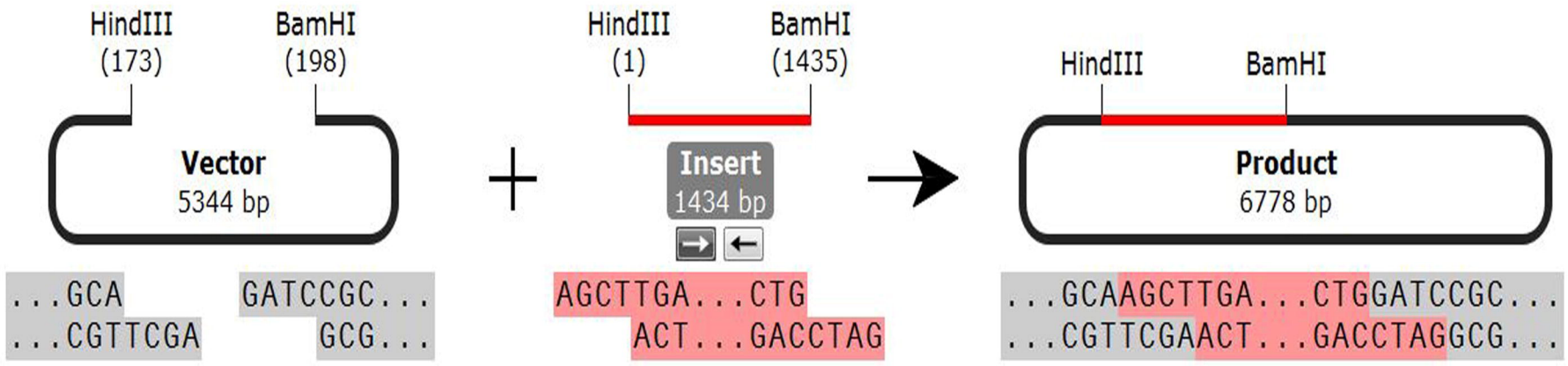
D


Residue index
Residue index

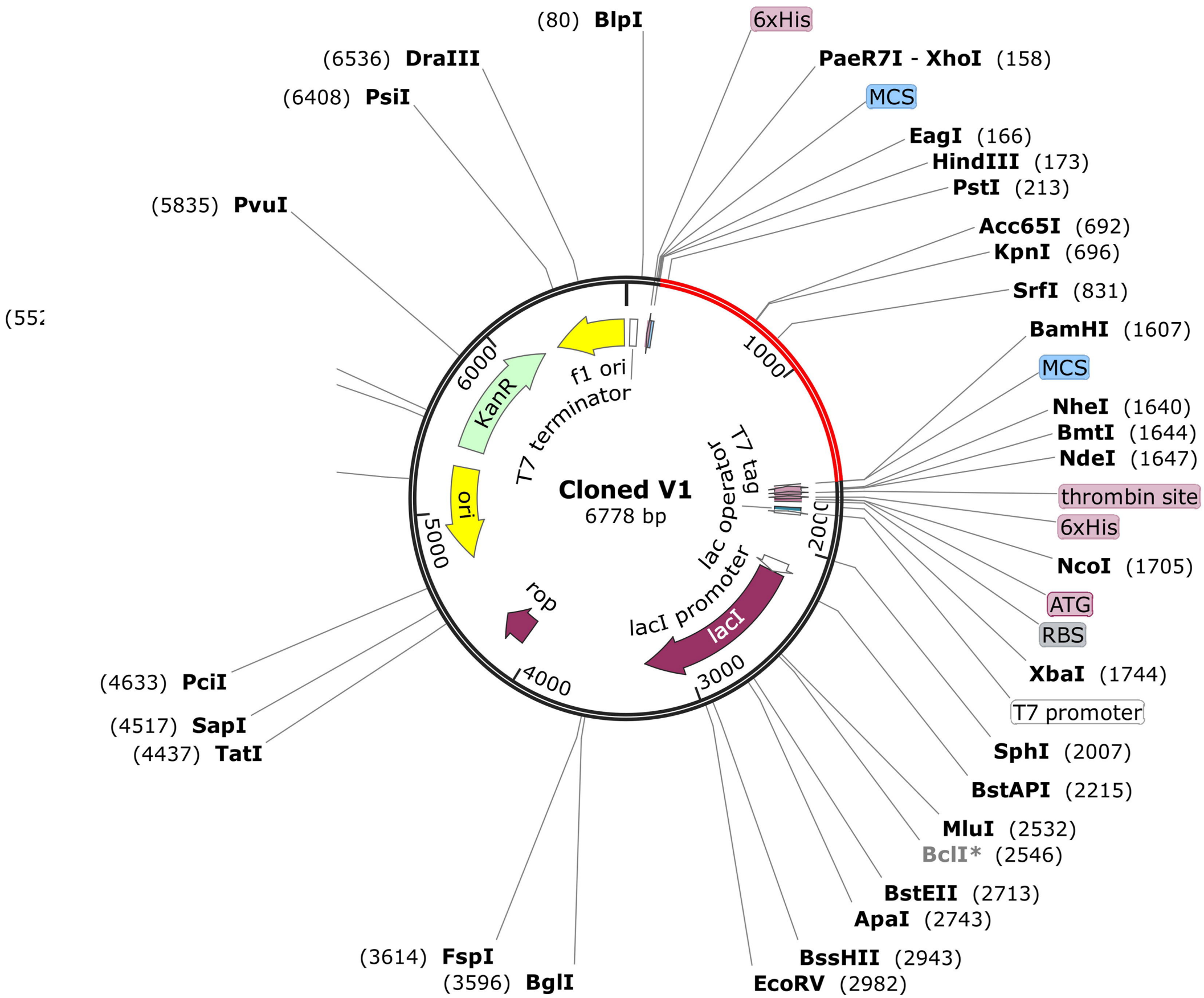
E


Atom index
Atom index

F


HindIII
(173)
BamHI
(198)
Vector
5344 bp
...GCA
...CGTTCGA
GATCCGC...
GCG...
HindIII
(1)
BamHI
(1435)
Insert
1434 bp
AGCTTGA...CTG
ACT...GACCTAG
HindIII
BamHI
Product
6778 bp
...GCAAGCTTGA...CTGGATCCGC...
...CGTTCGAACT...GACCTAGGCG...

A


(80) BlpI
6xHis
(6536) DraIII
PaeR7I - XhoI (158)
(6408) PsiI
MCS
EagI (166)
HindIII (173)
PstI (213)
(5835) PvuI
Acc65I (692)
KpnI (696)
SrfI (831)
BamHI (1607)
MCS
NheI (1640)
BmtI (1644)
NdeI (1647)
thrombin site
6xHis
NcoI (1705)
ATG
RBS
XbaI (1744)
T7 promoter
SphI (2007)
BstAPI (2215)
MluI (2532)
BclI* (2546)
BstEII (2713)
ApaI (2743)
BssHII (2943)
EcoRV (2982)
(3614) FspI
(3596) BglI
(4633) PciI
(4517) SapI
(4437) TatI
Cloned V1
6778 bp
KanR
f1 ori
T7 terminator
T7 tag
lac operator
lacI promoter
lacI
ori
rop
1000
2000
3000
4000
5000
6000

B