




Proteome Exploration of *Legionella pneumophila* To Identify Novel Therapeutics: a Hierarchical Subtractive Genomics and Reverse Vaccinology Approach

Md Tahsin Khan,^a Araf Mahmud,^a Mahmudul Hasan,^b Kazi Faizul Azim,^c Musammat Kulsuma Begum,^a Mohimenul Haque Rolin,^a Arzuba Akter,^d  Shakhinur Islam Mondal^a

^aDepartment of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet, Bangladesh

^bDepartment of Pharmaceuticals and Industrial Biotechnology, Sylhet Agricultural University, Sylhet, Bangladesh

^cDepartment of Microbial Biotechnology, Sylhet Agricultural University, Sylhet, Bangladesh

^dDepartment of Biochemistry and Molecular Biology, Shahjalal University of Science and Technology, Sylhet, Bangladesh

ABSTRACT *Legionella pneumophila* is the causative agent of a severe type of pneumonia (lung infection) called Legionnaires' disease. It is emerging as an antibiotic-resistant strain day by day. Hence, identifying novel drug targets and vaccine candidates is essential to fight against this pathogen. Here, attempts were taken through a subtractive genomics approach on the complete proteome of *L. pneumophila* to address the challenges of multidrug resistance. A total of 2,930 proteins from *L. pneumophila* proteome were investigated through diverse subtractive proteomics approaches, e.g., identification of human nonhomologous and pathogen-specific essential proteins, druggability and "anti-target" analysis, subcellular localization prediction, human microbiome nonhomology screening, and protein-protein interaction studies to find out effective drug and vaccine targets. Only three fulfilled these criteria and were proposed as novel drug targets against *L. pneumophila*. Furthermore, outer membrane protein TolB was identified as a potential vaccine target with a better antigenicity score. Antigenicity and transmembrane topology screening, allergenicity and toxicity assessment, population coverage analysis, and a 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 major histocompatibility complex (MHC) molecules and human immune TLR-2 receptors with minimum deformability at the molecular level. The present study aids the development of novel therapeutics and vaccine candidates for efficient treatment and prevention of *L. pneumophila* infections. However, further wet-lab-based phenotypic and genomic investigations and *in vivo* trials are highly recommended to validate our prediction experimentally.

IMPORTANCE *Legionella pneumophila* is a human pathogen distributed worldwide, causing Legionnaires' disease (LD), a severe form of pneumonia and respiratory tract infection. *L. pneumophila* is emerging as an antibiotic-resistant strain, and controlling LD is now difficult. Hence, developing novel drugs and vaccines against *L. pneumophila* is a major research priority. Here, the complete proteome of *L. pneumophila* was considered for subtractive genomics approaches to address the challenge of antimicrobial resistance. Our subtractive proteomics approach identified three potential drug targets that are promising for future application. Furthermore, a possible vaccine candidate, "outer membrane protein TolB," was proposed using reverse vaccinology analysis. The constructed vaccine candidate showed higher binding interaction with MHC molecules and human immune TLR-2 receptors at the molecular level. Overall, the present study aids in developing novel therapeutics and vaccine candidates for efficient treatment of the infections caused by *L. pneumophila*.

Editor Sébastien P. Faucher, McGill University

Copyright © 2022 Khan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Shakhinur Islam Mondal, shakhin200-gen@sust.edu, or Arzuba Akter, arzuba-bmb@sust.edu.

The authors declare no conflict of interest.

Received 30 January 2022

Accepted 24 June 2022

Published 12 July 2022

KEYWORDS *Legionella pneumophila*, Legionnaires' disease, molecular dynamics, protein-protein docking, reverse vaccinology, subtractive genomics

Legionella pneumophila is a human pathogen distributed worldwide within fresh-water and biofilms (1–3). These fastidious Gram-negative aerobic bacilli belong to the family *Legionellaceae* and the genus *Legionella*. The genus includes 60 other species and, together, these species are divided into 80 serogroups (4). Inhalation of *Legionella* spp. causes Legionnaires' disease (LD), a severe form of pneumonia and respiratory tract infection (5), while serogroup one alone is responsible for 70 to 90% of cases (6, 7). *L. pneumophila* is responsible for maximum incidences of water-associated diseases in the United States (8, 9) and caused more than half of all reported water-borne disease outbreaks in 2013 and 2014 (8), and the incidence rate of LD increased by 300% between 2005 and 2019 (10). It was declared an important pathogen by the U.S. Environmental Protection Agency (EPA) Contaminant Candidate List (CCL) due to its prevalence and the severity of its disease-causing effects (11). Diverse *Legionella* LD-causing species are genetically isolated from each other and selectively acquire genes with potential for enhanced virulence; however, they share a conserved core genome (12). Several virulence factors are encoded by distinct regions of DNA present in the pathogenicity island (PAI) of pathogenic bacteria (absent in nonpathogenic strains) (13–15). Moreover, *Legionella* possesses many of the traditional bacterial determinants that are important for pathogenicity in other bacteria, including pili, flagella, lipopolysaccharide (LPS), the type II secretion system, and several outer membrane proteins (OMPs), i.e., outer membrane vesicles, peptidoglycan-associated lipoprotein, major secreted phospholipase (PlaA), major cell-associated phospholipase A/lysophospholipase A (PlaB), major OMP, PlaC, macrophage infectivity potentiator, Hsp60, and FeoB (16). *Legionella* transmission from human to human generally does not occur or is rare and can only be acquired from environmental sources (17). The pathogen grows and localizes in hot, moist, and nutrient-rich sedimentary environments (18). A potable water distribution system was the culprit behind hospital-acquired Legionnaires' disease (19).

Macrolides (azithromycin) and fluoroquinolones (levofloxacin, moxifloxacin, and ciprofloxacin) are the standard treatment for *Legionella* infection (20, 21). Current American and European guidelines recommend azithromycin and levofloxacin as the first-line treatment (22). A combination of fluoroquinolone and azithromycin is recommended in the case of drug resistance or severe LD pneumonia (23). The emergence of resistance against antibiotics or a combination of antibiotics for various infectious agents poses global threats in the human health field (24). Recent epidemiological research found that *L. pneumophila* strains display a high prevalence of tolerance (50 to 100%) to widely used antibiotics such as azithromycin, ceftriaxone, rifampicin, tigecycline, ciprofloxacin, moxifloxacin, doxycycline, erythromycin, levofloxacin, and clarithromycin (25, 26). Despite the success of macrolides and fluoroquinolone, resistance developed against these drugs resulted in treatment failures (27, 28). Horizontal gene transfer is a driving force of acquiring antibiotic resistance genes, resulting in genome plasticity in bacteria (29). Natural transformation by competence plays a crucial role in horizontally acquired resistance in bacteria (30). Antibiotics of the fluoroquinolone family and genotoxic stress caused by UV radiation and other DNA-damaging agents induce competence genes in *L. pneumophila* resulting in rapid uptake and propagation of exogenous DNA (5, 31). Horizontal gene transfer by conjugative plasmids facilitates the incorporation of antibiotic resistance genes (32). Even transposons and bacteriophages are suitable for transferring antibiotic-resistant genes in pathogenic bacteria (33). *L. pneumophila* contains a type IV secretion system that makes it ideal for incorporating conjugative plasmids, leading to plasmid-mediated antibiotic resistance genes (30). LpeA, a subunit of the LpeAB efflux pump in *L. pneumophila* (homologous to AcrA in *Escherichia coli*), was attributed to macrolide resistance in the *L. pneumophila* pathogen (34). Researchers also detected point mutations or insertions/deletions in the

ribosomal genes *rplD* and *rplV* that increased the MICs of macrolides (27). Mutations in *gyrA*, *gyrB*, *parC*, and *parE* are associated with fluoroquinolone resistance (35). Therefore, there is an urgent need to develop new therapeutic antibacterial agents directed toward novel targets. A disease-based approach of the traditional drug discovery method is 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 time consumption by eliminating compounds in this process that have a limited chance of success (36–38). Among various computational methods and strategies, subtractive and comparative microbial genomics approaches efficiently identified potential targets in several human pathogens (39). The main theme of subtractive and comparative genomics approaches is to find targets (gene/protein) that are essential for the pathogen and possess no homology counterpart in the host (40). Along with the subtractive genomics approach, the 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*. Although successful in many cases, conventional or live-attenuated vaccines are often associated with several obstacles (41). Such vaccines may take decades to develop while relying on adequate antigen expression from *in vitro* culture models. Microorganisms are sometimes difficult to cultivate and, in some instances, attenuate (42, 43). Moreover, attenuation may cause cross-contamination and produce undesirable adverse immune responses (44). Epitope-based vaccine prediction, on the contrary, could be a faster approach to targeting immunogenic protein through the entire bacterial or viral proteome (45–47). The reverse vaccinology approach usually identifies a protein vaccine candidate using defined features, such as protein subcellular localization, topology, adhesion/antigenicity probability, epitopes, and binding to major histocompatibility complex (MHC) class I and II molecules (48). This approach has been proven to prioritize and design vaccine targets against multiple pathogens (49–52).

In this study, we utilized an in-depth subtractive genomics and reverse vaccinology approach to identify novel therapeutic drug and vaccine targets 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 nonhomologous to the host as well as host microbiota. We screened for outer membrane proteins (OMPs) and B- or T-cell epitopes. At the same time, we performed the human microbiome nonhomologous analysis. The highest scoring OMPs and epitopes will facilitate future *in vitro* and *in vivo* tests to produce drugs and suitable vaccine against intracellular *L. pneumophila* infection.

RESULTS

Genome-wide proteome exploration of *L. pneumophila* was employed to identify novel drug and vaccine targets through subtractive genomics and reverse vaccinology approaches. The subtractive genomics approach is a continuous process where irrelevant or nontarget sequences are allowed to be subtracted from a vast pool of initially retrieved sequences. Here, a list of human “anti-target” proteins was used to cross-check with the screened drug target alongside other common procedures of subtractive genomics. Moreover, the human microbiome was also checked with the screened putative drug candidates since the unintentional interference of proteins available in the gut microflora may cause adverse consequences on host physiology. After successfully screening sequences, we added reverse vaccinology approaches to get an insight into the potential design of vaccine candidates. During the entire procedure, the study utilized up-to-date bioinformatics approaches to discover novel therapeutics against *L. pneumophila*. The overall workflow has been illustrated in Fig. 1 and 2. All of the methodologies used in the proteome exploration of *L. pneumophila*, including the number of proteins screened in each step, are summarized in Table 1.

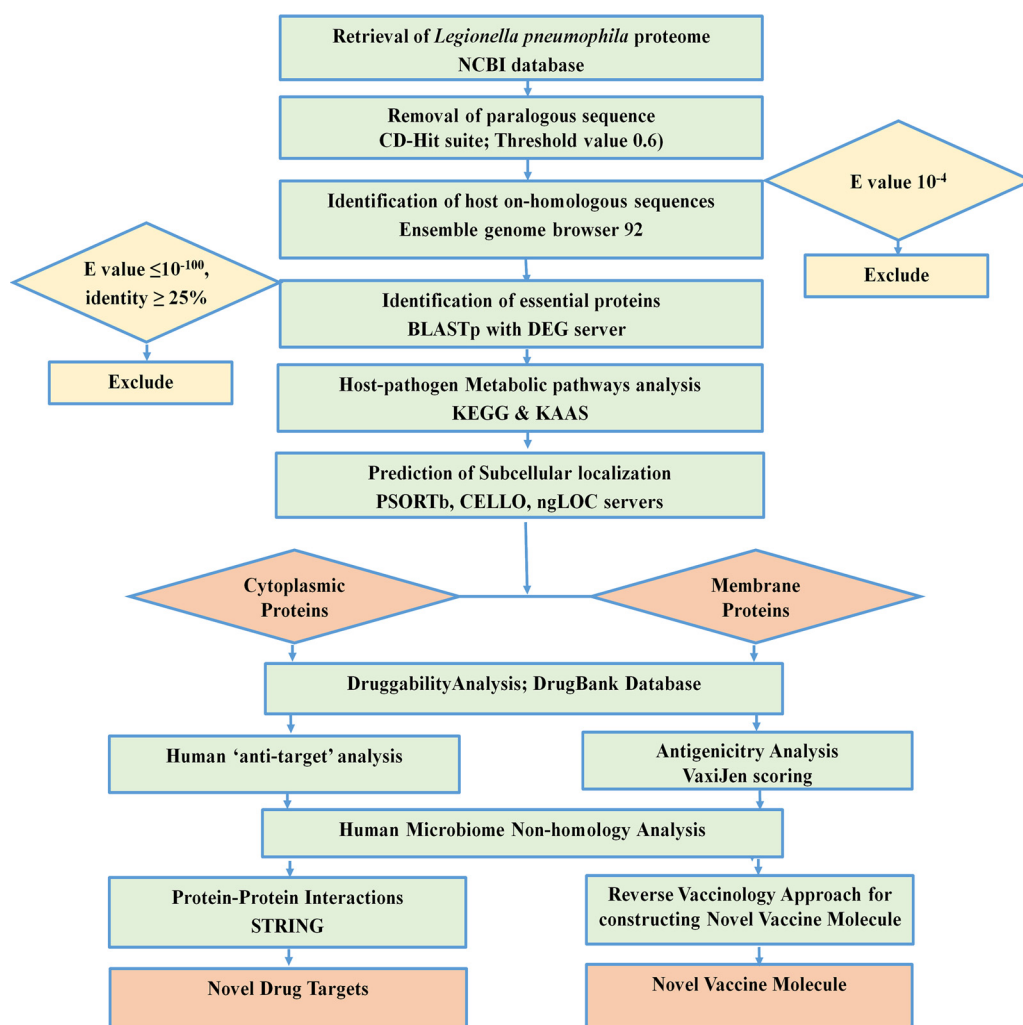


FIG 1 Proteome exploration of *L. pneumophila* to identify novel drug targets.

Proteome retrieval and finding paralogous sequences. The whole proteome retrieved from UniProt (Proteome ID UP000000609) containing 2,930 proteins was analyzed in the CD-Hit tool using the threshold value 0.6 to eliminate paralogous and duplicate sequences showing >60% sequence identity. Paralogous protein sequences were removed, leaving 2,878 nonparalogous protein sequences.

Identification of human nonhomologous proteins. BLASTp screened nonparalogous protein sequences through Ensembl genome browser 92. A total of 868 proteins were found to be similar to the human genome, which was excluded, leaving a total of 2,010 nonhomologous proteins.

Identification of essential proteins. Among the 2,010 nonhomologous proteins, only 125 protein sequences showed similarity in sequence identity with the essential proteins enlisted in the Database of Essential Genes (DEG) server from 48 bacterial strains. These 125 proteins are listed as essential proteins and are responsible for the survival of *L. pneumophila*.

Metabolic pathway analysis and family prediction of hypothetical proteins. Although 120 pathways of target *L. pneumophila* were identified from the KEGG server, only 37 were detected as unique for the pathogen (see Table S1 in the supplemental material). The result from the KAAS server at KEGG revealed that 122 proteins among 125 had assigned KEGG Orthology (KO) identifiers. Of the 122 proteins, only 4 were involved in unique metabolic pathways (Table 2), and 29 were involved in neither unique nor common pathways (see Table S2). Only three hypothetical proteins were found among these 29 KO-

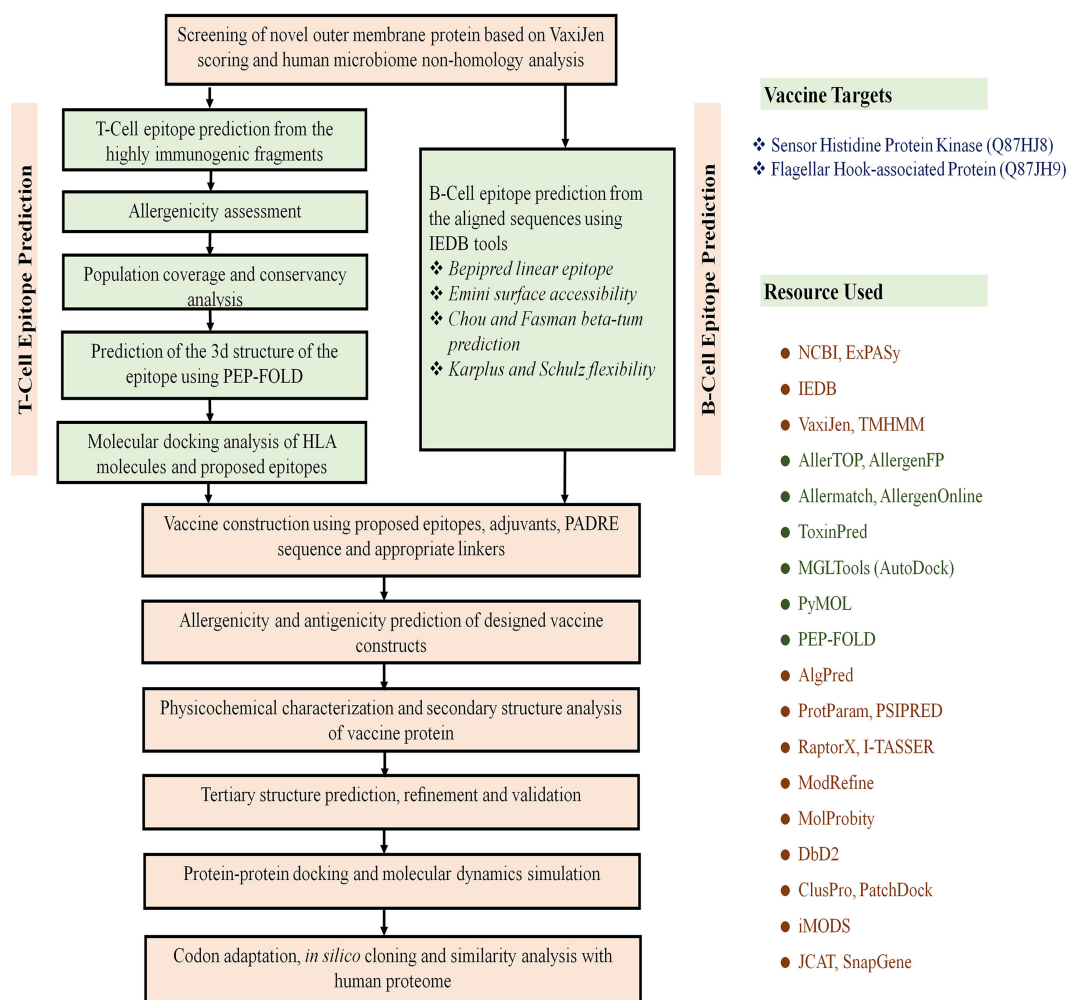


FIG 2 Flow chart summarizing the protocols for the prediction of epitope-based vaccine candidate by the *in silico* reverse vaccinology technique.

assigned proteins, and function of these three proteins remain unknown. Functional families for these hypothetical proteins were also analyzed (see Table S3). All of these 33 proteins were allowed for further investigation. The 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 of 16 proteins were identified as cytoplasmic proteins, and 16 were identified as membrane proteins which were further screened for druggability and antigenicity analysis (see Table S4). However, the study failed to predict the subcellular localization of the hypothetical protein (tr|Q5ZYU2) since these four servers generated different predictions.

Druggability screening and antigenicity analysis. Only seven proteins showed similarity with the available drug targets (Table 3) in the DrugBank 5.1.0 database. The remaining 25 showed no hits; these were considered novel drug targets of *L. pneumophila*, comprising of 13 cytoplasmic and 12 membrane proteins. Antigenicity analysis of membrane protein revealed that only one protein was nonantigenic, whereas the remaining 11 showed antigenicity scores of >0.4 according to the VaxiJen server (Table 4). Furthermore, all the proteins screened after druggability and antigenicity analysis were used for human “anti-target” and microbiome nonhomology studies.

“Anti-target” and human microbiome nonhomology analysis. Proteins that are important for human cellular homeostasis and trigger hazardous side effects under the

TABLE 1 Subtractive genomic analysis scheme toward the identification of novel therapeutic targets

Serial no.	Steps/description	<i>L. pneumophila</i> subsp. <i>pneumophila</i> ^a
1	Total no. of proteins	2930
2	No. of nonparalogous proteins (>60% identical) in CD-Hit	2878
3	No. of proteins nonhomologous to <i>H. sapiens</i> using BLASTp (E = 10 ⁻³)	2010
4	Essential proteins in DEG 15.2 server (E ≤ 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
	Protein of unknown location	1
10	Essential proteins found to be novel in DrugBank 5.1.0	25
11	Proteins nonhomologous to “anti-targets” using BLASTp (E < 0.005, identity <30%, query length >30%)	25
12	Membrane proteins showing antigenicity using VaxiJen v2.0 (threshold value >0.4)	12
13	No. of cytoplasmic and membrane proteins less similar to human microbiota using BLASTp against “43021[BioProject]” (E = 1, identity <47%)	5
14	Proteins selected for epitope design	1
15	Novel drug targets	3

^aStrain Philadelphia 1/ATCC 33152/DSM 7513.

influence of a drug are termed “anti-targets” (53). To avoid severe cross-reaction and toxic effects in humans, identifying nonhomologous proteins as human “anti-targets” is a crucial step. The study revealed that all selected 25 novel therapeutic proteins showed no similarity with 210 human “anti-targets.” In addition, BLASTp analysis of all microbial strains in the Human Microbiome Project (HMP) server carried out using the NCBI blast server revealed that only 5 of 25 proteins showed a similarity of ≤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. Among the five novel proteins with ≤47% similarity with human microbiota, three were suggested as novel drug targets for *L. pneumophila* based on protein-protein interaction (PPI) study (Table 6). PPIs in STRING v11.5 revealed that the cytoplasmic protein UDP-*N*-acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase (*murE*) interacted with proteins involved in cell wall biosynthesis pathways (Fig. 3A), whereas “trigger factor (*tig*)” had interactions with proteins that were mainly involved in binding with different rRNA and tRNA molecules (Fig. 3B). Moreover, the inner membrane protein “probable lipid II flippase MurJ (*mvjN*)” interacted with proteins that were connected to the peptidoglycan biosynthesis pathway and transportation of lipid-linked peptidoglycan precursors (Fig. 3C), where “biotin synthase (BioB)” is an essential protein for the survival of the pathogen. Important drug targets are listed in Table 6. Another inner membrane protein, “probable potassium transport system protein Kup1 (*kup1*),” presented a single interaction (Fig. 3D), and as a consequence, was excluded from our analysis. Moreover, the remaining outer membrane protein “TolB (*tolB*)” showed close interactions with other proteins (Fig. 3E), mainly found in peptidoglycan synthesis and lipoprotein translocation. Protein TolB also exhibited less similarity (<47%) with human microflora proteins and had an antigenicity

TABLE 2 Proteins involved in pathogen specific pathways

Serial no.	KO assignment	Accession no.	Protein name	Pathway
1	K18138	tr Q5ZXL1	Multidrug resistance protein	Beta-lactam resistance; cationic antimicrobial peptide (CAMP) resistance
2	K01928	tr Q5ZX16	UDP- <i>N</i> -acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase	Lysine biosynthesis; peptidoglycan biosynthesis
3	K03587	tr Q5ZX17	Peptidoglycan D,D-transpeptidase FtsI	Peptidoglycan biosynthesis; beta-lactam resistance
4	K05515	tr Q5ZVR5	Peptidoglycan D,D-transpeptidase MrdA	Peptidoglycan biosynthesis; beta-lactam resistance

TABLE 3 Identified druggable targets with drug names

Serial no.	Accession no.	Protein name	DrugBank ID	Drug name ^a	Bit score	Query length (aa) ^b
1	tr Q5ZS52	RNA polymerase sigma factor RpoH	DB08874	Fidaxomicin	103.605	318
2	tr Q5ZT04	RNA polymerase sigma factor RpoD	DB08226 and DB08266	Myxopyronin B and methyl [(1 <i>E</i> ,5 <i>R</i>)-5-{3-[(2 <i>E</i> ,4 <i>E</i>)-2,5-dimethyl-2,4-octadienoyl]-2,4-dioxo-3,4-dihydro-2 <i>H</i> -pyran-6-yl} hexylidene]carbamate	302.368	248
3	tr Q5ZUP0	UDP- <i>N</i> -acetylmuramate-L-alanyl-gamma-D-glutamyl-meso-2,6-diaminoheptandioate ligase	DB01673, DB03909, and DB04395	Uridine-5'-diphosphate- <i>N</i> -acetylmuramoyl-L-alanine, adenosine-5'-[β,γ-methylene] triphosphate, and phosphoaminophosphonic acid-adenylate ester	145.591	398
4	tr Q5ZXL1	Multidrug resistance protein	DB03825, DB04209, and DB03619	Rhodamine 6G, dequalinium, and deoxycholic acid	427.557	1,020
5	tr Q5ZX17	Peptidoglycan D,D-transpeptidase FtsI	DB00303 and DB00671	Ertapenem and cefixime	122.094	632
6	tr Q5ZVR5	Peptidoglycan D,D-transpeptidase MrdA	DB01598, DB01329, DB01327, DB01163, DB01328, DB01413, DB01415, DB00948, DB00438, DB00303, and DB06211	Imipenem, cefoperazone, cefazolin, amdinocillin, cefonicid, cefepime, ceftibuten, mezlocillin, ceftazidime, ertapenem, and doripenem	486.493	612
7	tr Q5ZZI4	Amino acid permease	DB00123, DB00125, and DB00129	L-Lysine, L-arginine, and ornithine	155.606	424

^aDrug names correspond respectively to the DrugBank ID numbers listed in Table 3, column 4.

^baa, amino acids.

score of >0.6, so it was considered a potential vaccine candidate against *L. pneumophila*. A further study was used for *in silico* vaccine design focusing on protein TolB by a reverse vaccinology approach. Details of PPIs are provided in Table S5 in the supplemental material.

Homology modeling, structure optimization, and validation. The selected single vaccine candidate, protein TolB (Q5ZV69), was modeled to determine three-dimensional (3D) structure through EasyModeller 4.0 (see Fig. S1A). Complete results of template selection for homology modeling are provided in Table S6. The best structure generated for each target was optimized by the ModRefiner server. The refined protein TolB (Q5ZV69) model had a root mean square deviation (RMSD) of 0.1444 and a tem-

TABLE 4 Probable antigenic proteins for vaccine targets

Serial 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 nonhomology analysis and subcellular localization

Serial no.	Accession no.	Protein factor	Microbiome similarity	Subcellular localization
1	tr Q5ZX16	MurE	<46	Cytoplasmic
2	sp Q5ZUD8	Tig	<45	Cytoplasmic
3	tr Q5ZS84	MviN	<47	Inner membrane
4	sp Q5ZW98	Kup1	<47	Outer membrane
5	sp Q5ZV69	TolB	<44	Inner membrane

plate modeling (TM) score of 0.9783 in the initial model. After refinement, the Protein Data Bank (PDB) structure of each protein was evaluated using PROCHECK and ERRAT. Psi-Phi pains in Ramachandran plots revealed that protein TolB (Q5ZV69) had 91.9% residues in the most favored regions, 7.8% in the additional allowed regions, and 0.3% in the generously allowed regions (see Fig. S1B). According to ERRAT, the overall quality factor of this protein was 76.1084 (see Fig. S2). An ERRAT score of 50 is usually acceptable, and the value explains the statistics of nonbonded 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, *Legionella pneumophila* (LEGPH) protein TolB (Q5ZV69) was selected based on PPI analysis with a total antigenicity score of 0.6740 for vaccine design against *Legionella*. The ProtParam tool was used to analyze the vaccine proteins' physicochemical properties (see Table S7). The molecular weight (45.38 kDa) of protein TolB indicated its good antigenic potential. At the same time, the theoretical pI of 8.67 showed that the protein would have a net positive charge below the pI and vice versa. The extinction coefficient was 34,380 at 0.1% absorption, assuming all cysteine residues are reduced. The GRAVY (grand average of hydropathy) value -0.204 described the hydrophilic nature of the protein, whereas the aliphatic index 89.81 ensured the thermostability of the protein. The instability index (38.09) classified the protein as a stable one. In addition, various immunogenic epitopes from LEGPH protein TolB were identified as potential T-cell epitopes that can bind to various HLA-A and HLA-B cells with greater 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 epitopes. The best LEGPH protein TolB epitopes were ranked based on the topology of the transmembrane and the antigenicity score (Table 7). All indicated alleles were identified as optimal binders of the proposed epitopes and used to assess population coverage (Fig. 4). Studies utilizing four separate allergenicity prediction systems have retained epitopes classified as nonallergenic to humans in the predicted epitope list (Table 7).

Toxicity and conservancy analysis of T-cell epitope. ToxinPred server predicted the relative toxicity of the top epitopes (Table 7). Again, putative epitopes generated from LEGPH protein TolB were highly conserved (up to 100%) within specific pathogenic strains. Top epitopes were used to design the final vaccine constructs to allow a broad spectrum of immunity (Table 7).

B-cell epitope identification and cluster analysis. B-cell epitopes of the selected protein were generated using a total of six algorithms (Table 8). For envelope glycoprotein, peptide sequences from residues 66 to 77 and residues 404 to 417 were identified as potential B-cell epitopes that could stimulate a preferred immune reaction (see

TABLE 6 Identified novel drug targets against *L. pneumophila*

Serial no.	Accession no.	Protein name
1	tr Q5ZX16	UDP- <i>N</i> -acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase (<i>murE</i>)
2	sp Q5ZUD8	Trigger factor (<i>tig</i>)
3	tr Q5ZS84	Probable lipid II flippase MurJ (<i>mviN</i>)

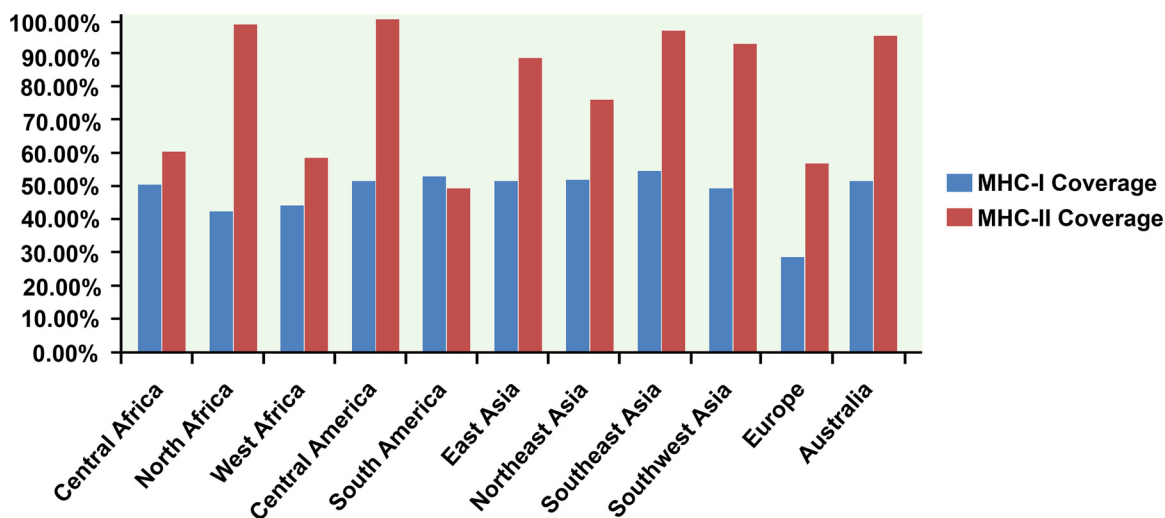


FIG 4 Population coverage analysis of LEGPH protein ToIB. Blue indicates MHC-I allele coverage, whereas red indicates MHC-II allele coverage.

314—were the most flexible (see Fig. S3D). Amino acids in the areas from residues 5 to 23 and from residues 243 to 252 were highly antigenic (see Fig. S3E), while the regions from residues 174 to 180 and from residues 283 to 289 were predominantly hydrophilic (see Fig. S3F). Before the final constructs were designed, peptides containing overlapping epitopes between the top T cells and B cells were identified (see Table S8).

Design of the vaccine construct. Both clusters and singletons (which contain a single epitope) were used to construct the final vaccine molecules. Three adjuvants—beta-defensin, ribosomal protein L7/L12, and *Mycobacterium tuberculosis* HABA protein (AGV15514.1)—were used to construct vaccines. The adjuvants were added at the N-terminal end of the vaccine construct, which helps the vaccine to create a more robust immune response in the patient’s body. The EAAAK, GGGG, GPGPG, and KK linkers were used to connect the epitopes in the correct position. The PADRE sequence was conjugated with the vaccine construct to ensure the maximum major histocompatibility complex class II (MHC-II) allele coverage. Three constructs with 476 (V1), 561 (V2), and 590 (V3) amino acid residues were designed and used for analysis (Table 9). A schematic representation of the constructed vaccine is shown in Fig. S4 in the supplemental material.

Allergenicity, antigenicity, physicochemical properties, and secondary structure analysis of vaccine constructs. The results revealed construct V1’s superiority due to a greater antigenicity score (0.603) and nonallergenic behavior (Table 9). The final construct of the vaccine was characterized by its physical and chemical properties.

TABLE 8 Allergenicity pattern of the predicted B-cell epitopes generated from LEGPH protein ToIB

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

^aaa, amino acids.

TABLE 9 Allergenicity prediction and antigenicity analysis of designed vaccine constructs

Vaccine construct	Composition	Complete sequence of vaccine construct	Allergenicity	VaxiJen score (threshold 0.4)
V1	Predicted CTL, HTL, and BCL epitopes of LEGPH protein TolB with β defensin adjuvant and PADRE sequence	EAAAKGIINTLQKYYCRVGRCAVLSCLPKKEEQIGKCSTRGR KCCRRKKEAAAKAFVAAWTLKAAAGGGSRIISLFLLLFTGGG STSGRGGSPQVYGGGSQVIALDLELTQGGGSIGVQNTGGGPG GGSLPAREGDVQGGGSDVDLSSGSMKGGGSTFGNSIDTEGPG PGTPKIYDVLSSGSMKQLTFGGPGPLTFGNSIDTEPRYSPDG GPGPGFTSGRGGSPQVYRLSLAGPGPGYSLEVADADGHNQPS GPGPGRVTFEGNYNARASYTGPGGGGRSRYLEVADADGGP GPGISGPQGANSQSSVSTGPGPGVIALDLELTQGINSAKKSGPQG ANSQSSVKKAREGDVQEPAWSPYKKSWSPNGKKKSGRGGSPKK SFEKKKAKKIDTEPRYSPDGKADGQISRKKIISLFLLLFTGQVIALDL EKKGQHLAVVLSKKKYS PDGRSKKAKFVAAWTLKAAAGGGS	Nonallergen	0.6034
V2	Predicted CTL, HTL, and BCL epitopes of LEGPH protein TolB with L7/L12 ribosomal protein adjuvant and PADRE sequence	EAAAKMAKLDDELDAFKEMTLLELSDVKKFEETFEVTAAPVA VAAAGAAPAGAAVEAAEEQSEFDVILEAAGDKKIGVIKVVREIVSGL GLKEAKDLVDGAPKPLLEKVAKEAEDA KAKLEAAGATVTVKEAAA KAKFVAAWTLKAAAGGGSRIISLFLLLFTGGGSTSGRGGSPQVYGGG SQVIALDLELTQGGGSIGVQNTGGGPGGSLPAREGDVQGGGSDVD LSSGSMKGGGSTFGNSIDTEGPGPGTPKIYDVLSSGSMKQLTFGGP GPGPLTFGNSIDTEPRYSPDGGPGFTSGRGGSPQVYRLSLAGPGPG YSLEVADADGHNQPSGPGPGRVTFEGNYNARASYTGPGGGGRSRY SLEVADADGGPGPGISGPQGANSQSSVSTGPGPGVIALDLELTQGINS AKKSGPQGANSQSSVKKAREGDVQEPAWSPYKKSWSPNGKKKSGR GGSPKKSFEKKKAKKIDTEPRYSPDGKADGQISRKKIISLFLLLFTGQV IALDLEKKGQHLAVVLSKKKYS PDGRSKKAKFVAAWTLKAAAGGGS	Nonallergen	0.5524
V3	Predicted CTL, HTL, and BCL epitopes of LEGPH protein TolB with HABA adjuvant and PADRE sequence	EAAAKMAENPIDDLAPLALGAADLALATVNDLIANLRER AEETRAETRTRVEERRARLTKFQEDLPEQFIELRDKFTTEELR KAAEGYLEAATNRYNELVERGEAALQRLRSQTAFEDASARAE GYVDQAVELTQEALGTVASQTRAVGERAAKLVGIELEAAAKAK FVAAWTLKAAAGGGSRIISLFLLLFTGGGSTSGRGGSPQVYGGGS QVIALDLELTQGGGSIGVQNTGGGPGGSLPAREGDVQGGGSDVD DLSSGSMKGGGSTFGNSIDTEGPGPGTPKIYDVLSSGSMKQLTF GGPGPLTFGNSIDTEPRYSPDGGPGFTSGRGGSPQVYRLSLA GPGPGYSLEVADADGHNQPSGPGPGRVTFEGNYNARASYTGPGP GGGRSRYLEVADADGGPGPGISGPQGANSQSSVSTGPGPGVIALD LELTQGINSAKKSGPQGANSQSSVKKAREGDVQEPAWSPYKKSWS PNGKKKSGRGGSPKKSFEKKKAKKIDTEPRYSPDGKADGQISRKKIISL LLFTGQVIALDLEKKGQHLAVVLSKKKYS PDGRSKKAKFVAAWTLK AAAGGGS	Nonallergen	0.5726

Construct V1’s molecular weight was 48.34 kDa, while the theoretical pI was measured at 9.78, indicating that the protein should have a net negative charge over the pI and vice versa. The half-life of the vaccine was expected to be more than 10 h in *Escherichia coli in vivo*. The estimated extinction rate and an aliphatic index were 41,370 and 61.74, respectively. The protein’s computed GRAVY value was -0.545 , while the instability index (33.51) classified the protein as stable. Constructed V1 was, in contrast, characterized by a 6.30% alpha-helix, a 24.78% sheet, and a 68.92% coil structure (see Fig. S5).

Vaccine tertiary structure prediction, refinement, validation, and disulfide engineering. 2YMU (chain A) from the PDB database was detected as the most suitable V1 template, and a single-domain 3D model was created through the Raptor X server (Fig. 5A). All 476 residues were modeled by the server, while only 8% of residues were in the disordered region. The 3D model’s *P* value was 5.30×10^{-11} , which ensured better-quality modeling of the proposed vaccine. Of the residues, 81.4, 11.8, and 6.8% were in the favored, allowed, and outlier regions, respectively, prior to refinement. However, 88.4% of the residues were in the favored region after refining. A Ramachandran plot showed that 8.4%- and 3.2% of the residues were in the allowed and outlier regions (Fig. 5B). Homology modeling of constructs V2 and V3 was also performed, as shown

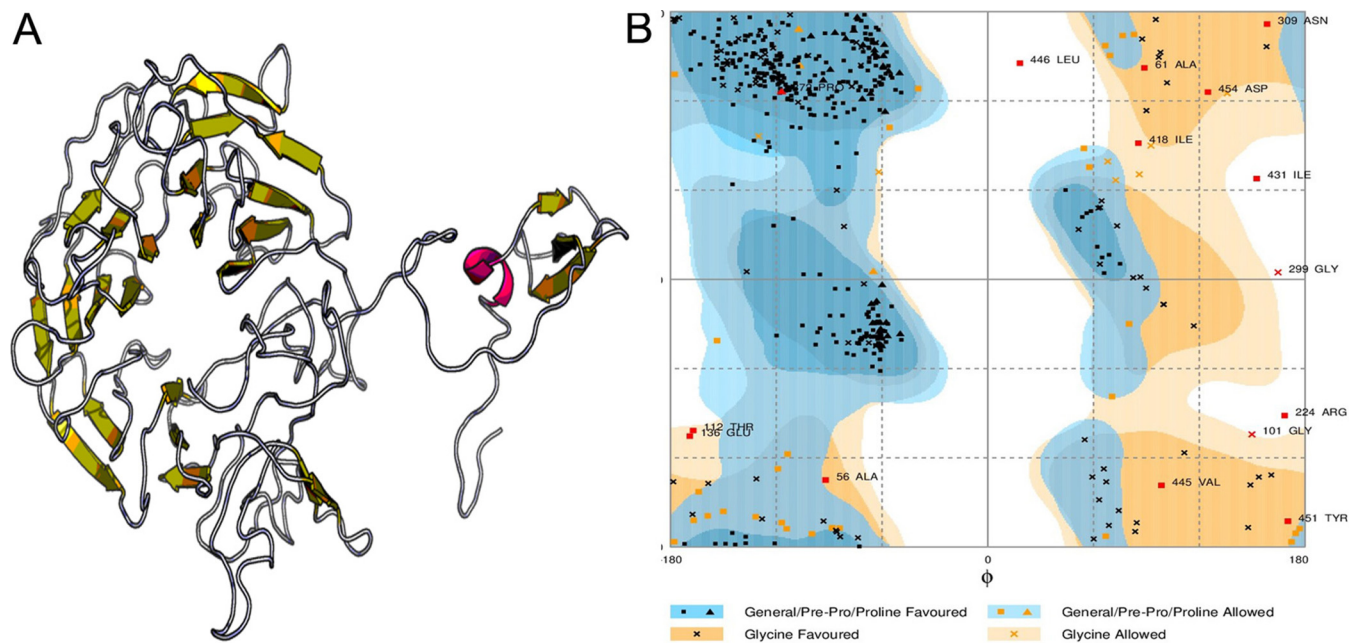


FIG 5 Tertiary structure prediction and validation of vaccine protein V1. (A) Structure diagram. (B) Validation of the 3D structure of vaccine protein V1 as determined by Ramachandran plot analysis.

in Fig. S6. For construct V1, 71 amino acid pairs could shape disulfide bonds. However, three pairs—ASN 202/GLY 226, SER 211/GLY 217, and ASN 309/LEU 317—based on the χ_3 and B-factor values met disulfide bond-forming requirements (Fig. 6).

Protein-protein docking, molecular dynamics simulation, and interaction analyses of vaccine constructs and TLR-2. Docking’s study revealed that construct V1 bound in the groove of different HLAs with minimal binding energy, which was biologically significant (Table 10). The predicted binding energies for the vaccine V1-TLR-2 complex were $-1,232.8$ and -275.55 kJ/mol as determined by ClusPro (Fig. 7) and the HDock server, respectively. FireDock output refinement of PatchDock server showed the lowest global energy of -15.64 for solution 2. The lowest binding energy measured the highest binding affinity between TLR-2 and vaccine construct. The molecular dynamics simulation of the vaccine onto the TLR-2 protein was performed for 100 ns.

The trajectory of the molecular dynamics (MD) simulation of the docked complex was analyzed to plot RMSD, root mean square fluctuation (RMSF), and radius of gyration (Rg) plots. RMSD plot (Fig. 8A) of the simulation trajectory calculated by referencing the backbone of the complex. The plot showed a minor deviation from the initial

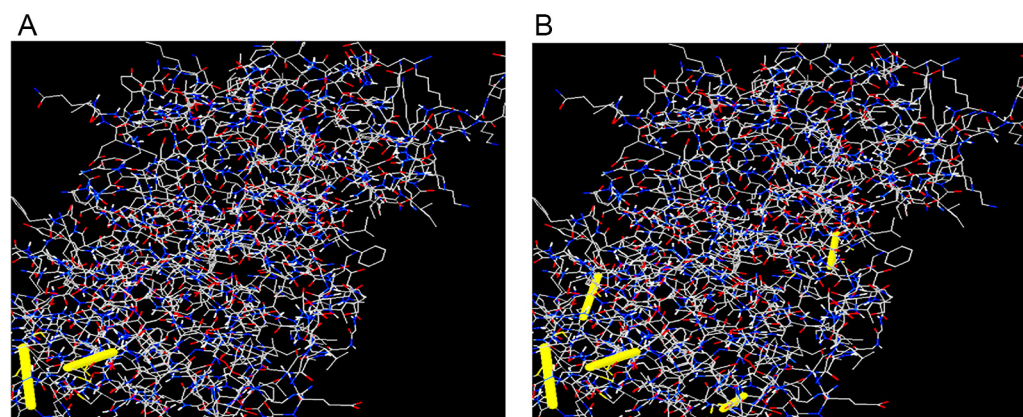


FIG 6 Disulfide engineering of vaccine protein V1. (A) Initial form (two disulfide bonds, indicated by yellow bands). (B) Mutant form (five disulfide bonds, indicated by yellow bands).

TABLE 10 Docking scores of vaccine construct V1 with different HLA alleles^a

HLA allele PDB ID	Global energy	Hydrogen bond energy	ACE	Score	Area
1A6A	-12.98	-4.93	7.18	17648	3,028.30
1H15	-30.47	-5.55	5.82	16608	2,414.80
2Q6W	-10.99	-2.58	7.65	17294	2,662.20
2SEB	-27.99	-8.91	0.83	18036	2,334.50
2FSE	-1.43	-0.46	0.05	19568	2,748.70
3C5J	-9.43	-3.63	15.90	16902	2,337.90

^aThese alleles included 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). ACE, atomic contact energy.

structure, indicating the docked complex's stability. Aside from the initial increase, in the beginning, the RMSD plot never crossed the 0.8 nm of deviation and stayed between 0.5 and 0.8 nm. The RMSF plot usually shows the fluctuations in the residues of a protein complex. The RMSF plot of the complex showed a change in the flexible regions of the receptor and minor fluctuation in the vaccine part (Fig. 8B). The Rg plot analyzes the compactness of a complex. The docked complex of the vaccine onto the receptor shows significant compactness suggesting adhesion onto the receptor (Fig. 8C). Docked complex (vaccine and chain B of TLR-2) produced seven hydrogen bond interactions and two salt bridges (Fig. 9; see also Table S9). These simulation and interaction analyses showed significant evidence of the stability of the receptor vaccine complex, suggesting that a proper immune response might occur for docking the vaccine onto the receptor complex.

Codon adaptation for *in silico* cloning and expression in *E. coli*. The optimized codon adaptation index was 0.95, and the GC content of that sequence was 54.01%. These results showed better expression in *E. coli* K-12. The optimized codons were introduced into vector pET28a(+) and the restriction sites HindIII and BamHI. A 6,346-bp clone was created [see Fig. S7; the desired 1,434-bp sequence of the pET28a(+) vector sequence is indicated in red].

DISCUSSION

Among all the pathogenic agents responsible for pneumonia and respiratory tract infection, *L. pneumophila* is one of the main causative agents showing high resistance against many antibiotics. Although clinical studies of antibiotic treatment against

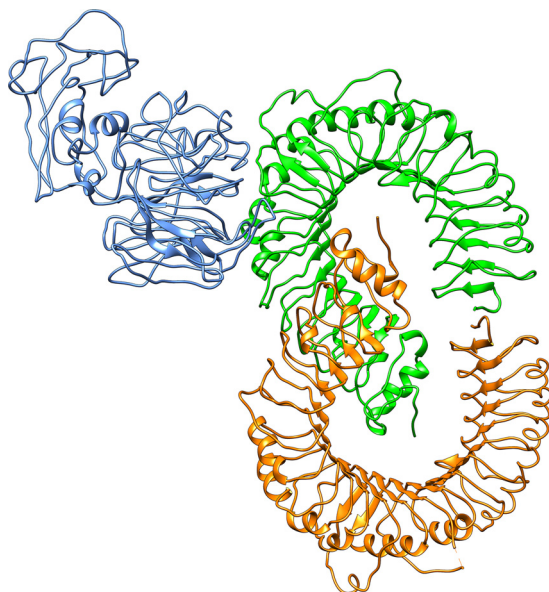


FIG 7 Docked complex of vaccine construct V1 with human TLR-2. Chain A of TLR-2 is indicated in orange, chain B of TLR-2 is indicated in green, and the vaccine molecule is indicated in blue.

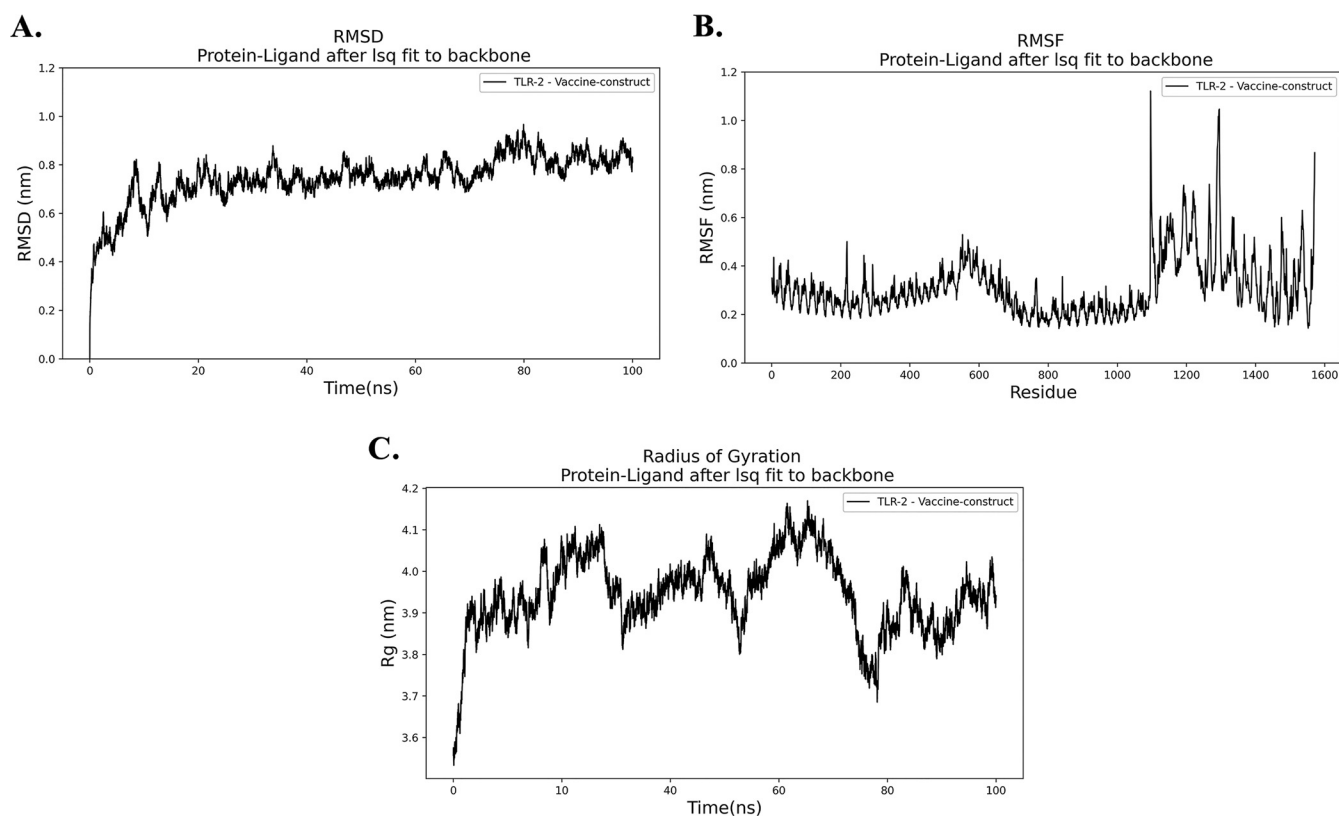


FIG 8 Molecular dynamics simulation of the vaccine protein V1/TLR-2 complex. (A) RMSD, root mean square deviation. (B) RMSF, root mean square fluctuation. (C) Rg, radius of gyration.

Legionnaires' disease have been reported, the number of such cases is insufficient (54). Hence, treatment recommendations are limited due to a lack of evidence (55). However, Rahimi and Vesal reported that *L. pneumophila* strains demonstrated the greatest resistance to ciprofloxacin, erythromycin, clarithromycin, moxifloxacin, and azithromycin (56). Therefore, the primary goal here was to screen out and determine novel drug targets and vaccine candidates for *L. pneumophila* through subtractive genomics and a reverse vaccinology approach.

Computational predictions help identify essential proteins for the pathogen's survival, nonhomologous to host proteins, and no involvement in the metabolic pathways of the host, leading to choosing the proteins only present and involved in pathogen-specific metabolic pathways. The complete reference proteome of *L. pneumophila* (2,930 proteins) was searched and downloaded from the NCBI protein database. Proteins involved in several common cellular systems emerged as homologous with the same functions between humans and bacteria during evolution (57, 58); these 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 (59) and can be considered pathogen-specific drug targets (60). The study revealed 125 unique, essential proteins of *L. pneumophila* that can be considered suitable drug targets. Localization is an important aspect of any possible drug target since the cellular functions of proteins are often localized in specific compartments of the cell; hence, studying subcellular localization will be helpful in understanding disease mechanisms, as well as in developing new vaccine candidates and drug targets. Although both membrane and cytoplasmic proteins could serve therapeutic targets (61), membrane proteins have been mostly reported for vaccine candidates (62). Hence, we used membrane proteins for vaccine construction, whereas cytoplasmic proteins were proposed as suitable drug targets. Again, the usage of antibiotics has already reduced their efficiency due to gene mutation and,

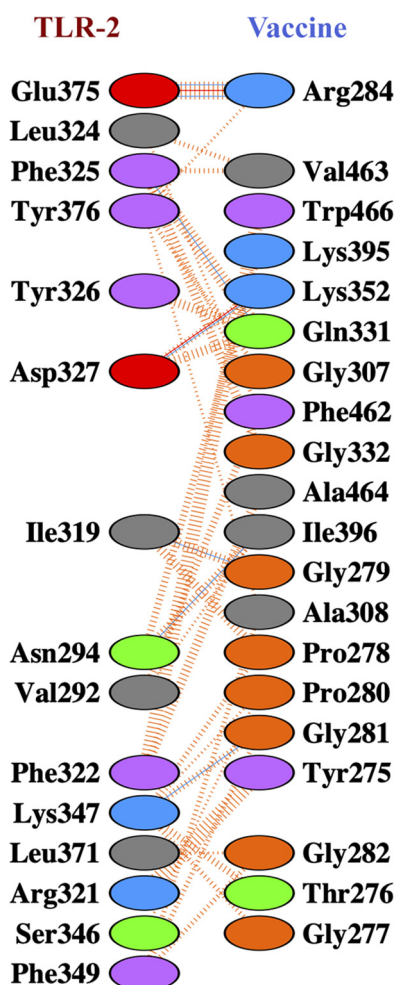


FIG 9 Vaccine construct V1-TLR-2 docked complex. Interacting residues are obtained after a 100-ns molecular dynamics simulation. Blue lines indicate hydrogen bonds, red lines indicate salt bridges and dashed orange lines indicate nonbonding interaction between the V1-TLR-2 complex.

thereby, the rapid emergence of resistant bacteria (63). Antibiotic and antimicrobial resistance crises have been observed due to the misuse and overuse of these drugs and to the scarcity of developing new drugs by the pharmaceutical industry (61, 64). Broad-spectrum drugs medicated for a pathogen or group of the pathogen may cause mutational changes and enhance the transfer of genes to other pathogens, which can show resistance to drugs, leading to the emergence of resistant bacteria. To avoid these problems, the selection of novel drug targets and vaccine candidates is a must; therefore, shortlisted 32 proteins were further screened through the DrugBank 5.1.0 database with default parameters.

Antigenicity scores for selected 12 novel drug-targeted membrane proteins revealed that only one protein was nonantigenic, whereas the remaining 11 showed antigenicity scores of >0.4 . These 11 proteins can be used to design B-cell and T-cell epitopes (specific for CTL and HTL) in the future to lessen the disease caused by *L. pneumophila*. Several drugs from different categories, such as nonnarcotic analgesics, nonsteroidal anti-inflammatory drugs, antidepressants, and vasodilators, were withdrawn from 1960 to 1999 due to various health-related cross-reactivity causing hepatic, cardiovascular, dermatological, and hematologic toxicities, as well as due to carcinogenic effects (65). For example, bromfenac (66), ebrotidine (67), and trovafloxacin (68) 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-reactions and adverse effects in humans, identifying nonhomologous proteins as human “anti-targets” (also referred to as essential proteins) was a crucial step in this study. Targeting human microbiome nonhomology proteins will be suitable since they are neither involved in common host-pathogen pathways nor homologous to any human “anti-targets.” Furthermore, drugs or vaccines 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 cytoplasmic proteins are involved in cell wall biosynthesis pathways and are essential for cell division. However, three proteins—UDP-*N*-acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase, trigger factor, and probable lipid II flippase MurJ—influence other crucial interacting proteins of the pathogen (Fig. 3) and were thus identified as the most attractive drug targets that can lead to the development of suitable therapeutics in the future. This study’s predicted single vaccine candidate, TolB (Q5ZV69), was analyzed to develop a potential and highly immunogenic vaccine candidate against *L. pneumophila*. Tol proteins are crucial for the survival and pathogenicity of many bacterial pathogens, such as *Escherichia coli*, *Haemophilus ducreyi*, *Vibrio cholerae*, and *Salmonella enterica* (69, 70). Protein TolB is involved in the translocation of the toxin across the outer membrane (69, 71). TolB is crucial in the Tol-Pal system since it directly interacts with Pal and other outer membrane proteins required for bacterial pathogenesis (72, 73). Apart from colicin uptake, the Tol-Pal system has a role in cell growth and fitness. *tolB* is part of the Tol-Pal system found in most Gram-negative bacteria that contains *tolA*, *tolQ*, *tolR*, and *pal* (74). Tol-Pal mutants contain potential vaccine properties and have promising therapeutic potential against multidrug-resistant bacterial pathogens (75). It was observed that the *tolB* mutant induces IgG production and gives a higher immunity level than do *tolA* and *tolR* mutants (76). Therefore, therapeutic inactivation of the TolB protein may affect and disrupt the interactions and synchronization of TolA, TolQ, TolR, and Pal proteins. The utilization of whole protein as a vaccine candidate is associated with the risk of the emergence of virulence property. Subunit vaccines, in contrast, can overcome such problems and consist of only the antigenic part of the pathogen with the potential to induce protective immunity within the host. Various antigenic epitopes were generated using different bioinformatics tools, which were extensively investigated for immunogenicity, toxicity profile, allergenicity pattern, and conservancy analysis. The PADRE sequence reportedly decreased the polymorphisms of HLAs in different populations (77, 78). In previous *in vivo* research, the linkers improved the immunogenicity of the vaccines (79, 80). Therefore, here, these critical considerations were considered in developing the final vaccine model. Moreover, the results of docking analysis revealed the binding affinity of promiscuous epitopes with different HLA alleles. In addition, the interaction between construct V1 and Toll-like receptor 2 (TLR-2) was checked to demonstrate the used adjuvants’ efficacy. TLR-2 is a major mammalian TLR recognizing lipoproteins derived from bacteria, viruses, fungi, and parasites (81). Literature reports suggest that TLR-2-mediated recognition of *Legionella* LPS and subsequent chemokine-dependent cellular recruitment is a crucial host innate response in *L. pneumophila* pneumonia (82). Therefore, we chose TLR-2 to evaluate its binding interaction with protein TolB, and our result revealed higher binding affinity, which is biologically significant. The vaccine-receptor complex also showed minimum deformability as well at the molecular level. Human *Legionella* infection has been studied in different animal models, most notably in mice and guinea pigs (83, 84). However, it was observed that the majority of mouse strains are resistant to *Legionella* infection and disease progression (85). Although the guinea pig model response to *Legionella* is closely similar to human infection, the scarcity of mutants and the high cost of specific immunological reagents discourage its use (86). Other animal models, such as *Caenorhabditis elegans* (87), *Drosophila melanogaster* (88), wax moth (89), and amoebae (90), have been developed to study *Legionella* infection. However, these models are not popular for different limitations (87, 88). Hence, the weakness of the existing models is a challenge to validate the vaccine target.

In conclusion, by using a subtractive genomics and reverse vaccinology approach, we are trying to develop novel therapeutics against *L. pneumophila* and may help to

reduce the rate of both mortality and morbidity caused by this pathogen. We mainly considered the key that is a must for survival of the pathogen, nonhomologous to the human, as well as human microbiota, and screened out B- and T-cell epitopes of OMPs. The highest scoring OMPs and epitopes will facilitate future wet lab-based experiments to develop drugs and suitable vaccine candidates against intracellular *L. pneumophila* infection. After intensive analysis, only three proteins were identified and proposed as novel therapeutic targets against *L. pneumophila*. Only OMP ToIB was identified as a potential vaccine candidate with a better antigenicity score. However, we suggest further *in vitro* and *in vivo* laboratory trials to validate our prediction.

MATERIALS AND METHODS

Proteome retrieval and finding paralogous sequences. The entire proteome of *L. pneumophila* subsp. *pneumophila* (strain Philadelphia 1/ATCC 33152/DSM 7513) (Proteome ID UP000000609), was retrieved from UniProt containing total 2930 proteins (91) which were subjected to CD-Hit analysis (<http://weizhong-lab.ucsd.edu/cdhit-web-server/cgi-bin/index.cgi>) (92). The sequence identity cutoff score was set at 0.6 to exclude redundant sequences of >60% identity through this method. Only nonparalogous protein sequences, excluding the redundant sequences (paralogous), were passed through further study.

Identification of human nonhomologous proteins. The goal of the nonhomology analysis of *L. pneumophila* subsp. *pneumophila* is to identify pathogen-specific proteins that are nonhomologous to the human. BLASTp analysis was carried out on nonparalogous proteins against the reference proteome of *Homo sapiens* using a threshold expectation value (E value) of 10^{-3} through Ensembl (<https://uswest.ensembl.org/Multi/Tools/Blast?db=core>) (93). Proteins that showed significant hits were filtered out, meaning they had similarities with the human genome leaving just the nonhomologous sequences. These nonhomologous sequences were screened. This step aims to avoid any functional resemblance with human proteomes and reduce any unwanted cross-reactivity of the drugs, thereby preventing the binding of drugs to the active site of homologous human proteins (94).

Identification of essential nonhomologous protein. Nonhomologous proteins were subjected to the DEG (59). DEG contains 53,885 essential genes and 786 essential noncoding sequences. BLASTp of previously selected nonhomologous proteins was carried out by selecting all the organisms present in DEG 15.2 server and using a threshold expectation value 10^{-10} and a minimum bit-score cutoff of 100 as parameters, but the hits with a E value of $\leq 10^{-100}$, an identity of $\geq 25\%$, and the same annotated function of the query were selected as essential proteins. These essential proteins were further subjected to metabolic pathway analysis.

Metabolic pathway analysis and family prediction of hypothetical proteins. The Kyoto Encyclopedia of Genes and Genomes (KEGG) includes complete metabolic pathways in living organisms (95). Using the three-letter KEGG organism codes "has" and "lpn" for human and *L. pneumophila* subsp. *pneumophila*, respectively, all of 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 only present in the pathogen as unique to *L. pneumophila* subsp. *pneumophila* using the KEGG pathway database (96), whereas the remaining pathways were grouped as common. The predicted human nonhomologous, essential proteins of *L. pneumophila* subsp. *pneumophila* were screened by BLASTp through KAAS (KEGG Automatic Annotation Server) (97) at KEGG to identify potential drug and vaccine targets. KAAS generates functional annotation of genes by BLAST comparisons against the manually compiled KEGG GENES database, and metabolic proteins are listed by assignments of KO (KEGG Orthology). It automatically generates KEGG pathways that indicate specific metabolic proteins. For the next phase of the subcellular localization study, proteins involved in these particular metabolic pathways of the pathogen and proteins allocated to KO but not involved in specific pathways were identified, except for proteins involved in common human and pathogen pathways. SVMProt is a server that uses a support vector machine to identify a protein sequence from its primary sequence into a specific class that contains all major groups of enzymes, transporters, receptors, pathways, and RNA- and DNA-binding proteins. The SVMProt server (98) was used to determine the functional classes of the pathogen's specific metabolic proteins in all uncharacterized, hypothesized proteins.

Prediction of subcellular localization. Because of the Gram-negative cell wall structure, *L. pneumophila* subsp. *pneumophila* proteins can be identified in five viable subcellular locations: cytoplasm, inner membrane, periplasm, outer membrane, and extracellular. Cytoplasmic proteins can be used as targets for medications, while surface membrane proteins can be used for both the medication and vaccine (99). PSORTb v3.0.2 (<http://www.psorb.org/psorb/index.html>) (100), CELLO v.2.5 (<http://cello.life.nctu.edu.tw/>) (101), and ngLOC (102) were used to predict the locations of selected nonhomologous essential pathogen proteins. The best score for a location was counted and generated by these servers. The final locations of these proteins were assessed via three steps: (i) if all three servers predicted the same location for a protein, that location was selected as the final result; (ii) if any two of these three servers predicted the same location for a protein, that location was selected as the final result; and (iii) if they predicted three different locations for a protein then PSLpred server (<https://webs.iitd.edu.in/raghava/pslpred/index.html>) (103) was used and the predicted location for that protein was matched with previous results predicted by PSORTb v3.0.2, CELLO v.2.5, and ngLOC servers.

Druggability screening and antigenicity analysis. DrugBank Database 5.1.0 (104) contains 2,556 approved small molecule drugs, 962 licensed pharmaceutical drugs, 112 nutraceuticals, and more than 5,125 experimental drugs. In addition, 5,117 nonredundant protein sequences (i.e., drug target/enzyme/

carrier/transporter) are linked to these drug entries. A “druggable” target must potentially bind to drugs and drug-like molecules with high affinity. Shortlisted cytoplasmic and membrane proteins were screened through the database of DrugBank 5.1.0 using default parameters. The drug targets from the shortlisted cytoplasmic and membrane proteins in the DrugBank database indicate that the same biological functions can act as evidence for their druggable properties and be grouped as current therapeutic targets. In contrast, their absence indicated those drug targets’ novelty and thus are classified as “novel therapeutic targets”. A total of 45 proteins that showed no similarity after passing through the DrugBank database were listed as novel drug targets and vaccine candidates. VaxiJen v2.0 (105) was used to predict protective antigens and subunit vaccines. Only membrane protein sequences of novel *L. pneumophila* subsp. *pneumophila* drug targets were subjected to VaxiJen v2.0, selecting a threshold value of 0.4. Proteins that showed antigenicity prediction >0.4 in VaxiJen v2.0 were transferred for the next step as potential vaccine candidates.

“Anti-target” analysis of the essential, nonhomologous, and novel drug targets. For humans, the human ether-a-go-go-related gene (hERG), constitutive androstane receptor (CAR), pregnane X receptor (PXR), and P-glycoproteins (P-gp) are anti-targets or alternative drug targets for host protein candidates. Some receptive membranes are also classified as “anti-target.” They are adrenergic α 1a, dopaminergic D2, muscarinic M1, and serotonergic 5-HT2C. A total of 210 human “anti-targets” have been identified in the literature (106), and the corresponding sequences of these proteins have been obtained from the NCBI protein database. BLASTp analysis was carried out for all nonhomologous, essential “novel drug target” proteins against these “anti-targets,” setting an E value threshold of <0.005 , a query length of $>30\%$, and an identity of $<30\%$ as parameters. Nonhomologous proteins showing $<30\%$ identity against these “anti-target” proteins were transferred to the next step.

Human microbiome nonhomology analysis. The correlation between gut flora and humans is not merely commensal but a symbiotic, mutualistic relationship (107), and different beneficial functions of the human microbiome were also reported (108, 109). Unintentional blocking or accidental inhibition of proteins present in this microflora may lead to adverse effects in the host (110). Screening of nonhomologous, essential proteins selected as vaccine candidates and novel drug targets were subjected to BLASTp through the NCBI protein blast server (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) using an E value cutoff score of 1 against the data set present in Human Microbiome Project server (<https://www.hmpdacc.org/hmp/>) “43021 (BioProject)” (111). The Human Microbiome Project (HMP) collected microbial strains from a disease-free, stable, 239-person adult population, which included 18 body habitats in five areas (oral, skin, nasal, gut, and urogenital), creating 5,026 microbial species compositions (112). Proteins showing a $<48\%$ similarity were identified as novel therapeutic targets and vaccine candidates and moved to the next stage. The results of this screening study ensured the chance to prevent unintended inhibition and unconscious blockage of human microflora proteins.

Protein-protein interaction studies. PPI analysis for selected shortlisted proteins were predicted using STRING (113). The database identifies physical (direct) and functional (indirect) interactions. PPI studies with a high confidence score ($\geq 70\%$) were considered to avoid false-positive results. Inhibition of essential proteins can hamper other proteins from performing correctly. Therefore, the PPI studies of the shortlisted proteins with others of the same strain can help identify the best therapeutic targets. Proteins showing close interactions with at least three others were selected for further studies. Only an OMP with antigenicity score of >0.6 was transferred to the next stage for reverse vaccinology after homology modeling and validation. The rest were listed as probable novel drug targets.

Homology modeling, structure optimization, and validation. Since no exact Protein Data Bank (PDB) structure was available for the selected proteins, a BLASTp analysis was carried out in the NCBI database for each protein, where PDB proteins were chosen as the database. The templates were chosen for homology modeling considering a sequence identity of ≥ 85 and a query cover of ≥ 30 . EasyModeller 4.0 software generated the best 3D structure of each selected therapeutic protein (114). The best models were subjected to ModRefiner for energy minimization and structure refinement (115). RMSD values were also calculated. Evaluation of the optimized, refined structures was carried out using the online servers PROCHECK (116) and ERRAT (117) present in the Structural Analysis and Verification Server (SAVES; <https://saves.mbi.ucla.edu/>) to get the stereochemical quality of the model.

Antigenic protein selection and T-cell epitope prediction. According to PPI analysis and antigenicity, the final vaccine target was selected among the membrane-associated proteins. The ProtParam (118) tool showed the different physicochemical parameters of the proteins. The IEDB MHC-I and MHC-II prediction tools predict MHC-I and MHC-II binding peptides, respectively (119).

Transmembrane topology, antigenicity, population coverage, and allergenicity assessment of T-cell epitopes. The TMHMM server predicted transmembrane helices in the protein (120), while the VaxiJen v2.0 server was used to test antigenicities of the epitopes (105). Population coverage has been evaluated for each human epitope using the IEDB analysis tool (121). Four independent servers—AllerTOP (122), AllergenFP (123), PA3P (124), and Allermatch (125)—are used to estimate epitope allergenicity for potential vaccine design.

Toxicity and conservancy analysis. The server ToxinPred (<https://webs.iitd.edu.in/raghava/toxinpred2/>) (126) estimated the relative toxicities of top T-cell epitopes. Conservancy of the epitope has been shown to determine the degree of distribution in the homologous protein set of corresponding epitopes using the IEDB epitope conservancy analysis tool (<http://tools.iedb.org/conservancy/>) (127).

Identification of B-cell epitopes and cluster analysis. Six different IEDB algorithms—the Kolaskar and Tongaonkar antigenicity scale (128), the Karplus and Schulz versatility prediction (129), the Chou and Fasman beta-turn prediction (130), the Emini surface accessibility prediction (131), the Parker hydrophilicity prediction (132), and the Bepipred linear epitope prediction tool (133)—were used to identify

the most active B-cell epitopes in the target protein. In addition, an analysis tool for the IEDB epitope cluster was used to identify overlapping peptides among the predicted epitopes (134). Larger cassettes containing multiple epitopes were found between top epitopes of CTL, HTL, and BCL.

Final vaccine construction. The final vaccine was formulated using the top epitopes (T and B cells), the adjuvants, and the corresponding linkers. Three specific adjuvants, i.e., beta-defensin, ribosomal protein L7/L12, and *Mycobacterium tuberculosis* HABA protein (AGV15514.1), have produced three separate vaccine molecules. Adjuvants interact with TLRs and cause immune activation by polarizing CTL reactions (135). The beta-defensin adjuvants are TLR-1, -2, and -4 agonists, whereas L7/L12 ribosomal protein and HBHA protein act only as TLR-4 agonists. EAAAK, GGGG, GPGPG, and KK linkers have been used to bind the adjuvant, CTL, HTL, and B-cell epitopes, respectively. Literature reports show that linkers allow efficient *in vivo* separation of individual epitopes (136, 137). To address the problem with highly polymorphic HLAs, a PADRE sequence was also inserted into vaccine constructs.

Allergenicity, antigenicity, physicochemical property, and secondary structure analysis of vaccine constructs. AlgPred v.2.0 (138) server was used to predict the nonallergic nature of the constructed vaccines. VaxiJen v2.0 server field (68) was used to suggest the superior vaccine candidate for evaluating its antigenicity. ProtParam (118) reveals the molecular weight, instability index, approximate half-life, isoelectric pH, GRAVY values, hydrophobicity, aliphatic index, and other physicochemical properties of the constructs. The secondary structure of the vaccine protein was determined by PSIPRED v3.3 (139) and NetTurnP v1.0 (140).

Vaccine tertiary structure prediction, refinement, validation, and disulfide engineering. The RaptorX server generated a tertiary structure of the constructs based on the degree of similarity between the target protein and accessible template structure in the PDB database (141, 142). Refinement was conducted using ModRefiner (115), followed by the FG-MD refinement server (143), to improve the accuracy of the predicted 3D modeled structure. A Ramachandran plot of the refined structure was assessed by RAMPAGE (144). Disulfide bonds enhance the geometric conformation of proteins and provide significant stability. DbD2 was used to build these bonds for the vaccine designs (145). The values of chi3 selected for the residue screening were -87 to +97, while the energy value considered was <2.5.

Protein-protein docking, molecular dynamics simulation, and interaction analysis of vaccine constructs and TLR-2. Inflammations caused by bacterial antigen are involved with TLR-2 immune receptors present over the immune cells (5, 146, 147). The 3D structure of different MHC molecules and human TLR-2s has been obtained from the database of RCSB proteins. A molecular docking method using ClusPro (148), HDOCK (149), and PatchDock server (150) determined the binding affinities of designed vaccines with specific HLAs and TLR-2 immune receptors. The FireDock server refined the complexes generated via the PatchDock server (150). The vaccine-receptor complex's stability was determined by contrasting the critical protein dynamics to their normal modes (151, 152). Essential dynamics is a versatile tool and an inexpensive solution to the costly atomistic simulation (151, 153).

Molecular dynamics simulation is an important step for analyzing the stability of a docked complex. Here, we evaluate the docked complex of TLR-2 and our V1 vaccine construct for stability. MD simulation was performed using the GROMACS 2020.1 package (154) using the CHARMM36 (155) force-field. The step was done to simulate the vaccine interaction with the TLR-2 in the native environment of the body. The docked complex of receptor-vaccine was solvated in a cubic box with SPC water molecules (156). To neutralize the system, 18 Cl⁻ ions were added to the system. The system underwent subsequent NVT and NPT ensemble MD simulations for 500 ps each time to maintain temperature and pressure. The system then underwent 100-ns MD simulation under equivalence at 300K and 1 bar. The simulation trajectory was then analyzed to plot the RMSD, RMSF, and Rg results. Finally, molecular interactions of the vaccine-TLR complex were analyzed using the PDBsum webserver to gain structural insights into the docked complex (157).

Codon adaptation and *in silico* cloning. A codon adaptation technique was used for greater expression of the vaccine protein in *E. coli*. The procedure was conducted by the JCAT server (158), while preventing the termination of rho's independent transcription, the ribosome-binding prokaryote site, and the cleavage of several other restriction enzymes (i.e., HindIII and BamHI). The optimized sequence of vaccine protein V1 was reversed and then conjugated with HindIII and BamHI restriction sites at the N- and C-terminal sites, respectively. SnapGene, a restriction cloning module, was used to insert the adapted sequence between HindIII (173) and BamHI (198) of the pET28a(+) vector.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 2.1 MB.

REFERENCES

1. Murga R, Forster TS, Brown E, Pruckler JM, Fields BS, Donlan RM. 2001. Role of biofilms in the survival of *Legionella pneumophila* in a model potable-water system. *Microbiology (Reading)* 147:3121–3126. <https://doi.org/10.1099/00221287-147-11-3121>.
2. Abu Khweek A, Amer AO. 2018. Factors mediating environmental biofilm formation by *Legionella pneumophila*. *Front Cell Infect Microbiol* 8:38. <https://doi.org/10.3389/fcimb.2018.00038>.
3. Steinert M, Hentschel U, Hacker J. 2002. *Legionella pneumophila*: an aquatic microbe goes astray. *FEMS Microbiol Rev* 26:149–162. <https://doi.org/10.1111/j.1574-6976.2002.tb00607.x>.
4. Miyashita N, Higa F, Aoki Y, Kikuchi T, Seki M, Tateda K, Maki N, Uchino K, Ogasawara K, Kiyota H, Watanabe A. 2020. Distribution of *Legionella* species and serogroups in patients with culture-confirmed *Legionella* pneumonia. *J Infect Chemother* 26:411–417. <https://doi.org/10.1016/j.jiac.2019.12.016>.

5. Newton HJ, Ang DK, Van Driel IR, Hartland EL. 2010. Molecular pathogenesis of infections caused by *Legionella pneumophila*. Clin Microbiol Rev 23:274–298. <https://doi.org/10.1128/CMR.00052-09>.
6. Yu VL, Plouffe JF, Pastoris MC, Stout JE, Schousboe M, Widmer A, Summersgill J, File T, Heath CM, Paterson DL, Cheresky A. 2002. Distribution of *Legionella* species and serogroups isolated by culture in patients with sporadic community-acquired legionellosis: an international collaborative survey. J Infect Dis 186:127–128. <https://doi.org/10.1086/341087>.
7. Fields BS, Benson RF, Besser RE. 2002. *Legionella* and Legionnaires' disease: 25 years of investigation. Clin Microbiol Rev 15:506–526. <https://doi.org/10.1128/CMR.15.3.506-526.2002>.
8. McClung RP, Roth DM, Vigar M, Roberts VA, Kahler AM, Cooley LA, Hilborn ED, Wade TJ, Fullerton KE, Yoder JS, Hill VR. 2017. Waterborne disease outbreaks associated with environmental and undetermined exposures to water—United States, 2013–2014. MMWR Morb Mortal Wkly Rep 66:1222–1225. <https://doi.org/10.15585/mmwr.mm6644a4>.
9. Beer KD, Gargano JW, Roberts VA, Hill VR, Garrison LE, Kutty PK, Hilborn ED, Wade TJ, Fullerton KE, Yoder JS. 2015. Surveillance for waterborne disease outbreaks associated with drinking water—United States, 2011–2012. MMWR Morb Mortal Wkly Rep 64:842–848. <https://doi.org/10.15585/mmwr.mm6431a2>.
10. CDC. 2018. National notifiable diseases surveillance system. National Notifiable Diseases Surveillance System, Centers for Disease Control and Prevention, Atlanta, GA.
11. Mraz AL, Weir MH. 2018. Knowledge to predict pathogens: *Legionella pneumophila* lifecycle critical review part I uptake into host cells. Water 10:132. <https://doi.org/10.3390/w10020132>.
12. Joseph SJ, Cox D, Wolff B, Morrison SS, Kozak-Muiznieks NA, Frace M, Didelot X, Castillo-Ramirez S, Winchell J, Read TD, Dean D. 2016. Dynamics of genome change among *Legionella* species. Sci Rep 6:33442–33414. <https://doi.org/10.1038/srep33442>.
13. Dowling JN, Saha A, Glew R. 1992. Virulence factors of the family *Legionellaceae*. Microbiol Rev 56:32–60. <https://doi.org/10.1128/mr.56.1.32-60.1992>.
14. Balasubramanian D, López-Pérez M, Grant T-A, Ogbunugafor CB, Almagro-Moreno S. 2022. Molecular mechanisms and drivers of pathogen emergence. Trends Microbiol <https://doi.org/10.1016/j.tim.2022.02.003>.
15. Meena M, Swapnil P, Zehra A, Aamir M, Dubey MK, Patel CB, Upadhyay R. 2019. Virulence factors and their associated genes in microbes, p 181–208. In *New and future developments in microbial biotechnology and bioengineering*. Elsevier, New York, NY.
16. Shevchuk O, Jäger J, Steinert M. 2011. Virulence properties of the *Legionella pneumophila* cell envelope. Front Microbiol 2:74. <https://doi.org/10.3389/fmicb.2011.00074>.
17. Mondino S, Schmidt S, Rolando M, Escoll P, Gomez-Valero L, Buchrieser C. 2020. Legionnaires' disease: state of the art knowledge of pathogenesis mechanisms of *Legionella*. Annu Rev Pathol 15:439–466. <https://doi.org/10.1146/annurev-pathmechdis-012419-032742>.
18. Simmering JE, Polgreen LA, Hornick DB, Sewell DK, Polgreen PM. 2017. Weather-dependent risk for Legionnaires' disease, United States. Emerg Infect Dis 23:1843–1851. <https://doi.org/10.3201/eid2311.170137>.
19. Sabria M, Yu VL. 2002. Hospital-acquired legionellosis: solutions for a preventable infection. Lancet Infect Dis 2:368–373. [https://doi.org/10.1016/s1473-3099\(02\)00291-8](https://doi.org/10.1016/s1473-3099(02)00291-8).
20. Wimer SM, Schoonover L, Garrison MW. 1998. Levofloxacin: a therapeutic review. Clin Ther 20:1049–1070. [https://doi.org/10.1016/s0149-2918\(98\)80104-5](https://doi.org/10.1016/s0149-2918(98)80104-5).
21. Brittain HG. 2020. Profiles of drug substances, excipients, and related methodology. Academic Press, New York, NY.
22. Woodhead M, Blasi F, Ewig S, Garau J, Huchon G, Ieven M, Orqvist A, Schaberg T, Torres A, van der Heijden G, Read R, Verheij TJM. 2011. Guidelines for the management of adult lower respiratory tract infections (full version). Clin Microbiol Infect 17:E1–E59. <https://doi.org/10.1111/j.1469-0691.2011.03672.x>.
23. Harris M, Clark J, Coote N, Fletcher P, Harnden A, McKean M, Thomson A. 2011. Guidelines of the management of community acquired pneumonia in children. British Thoracic Society Community Acquired Pneumonia in Children Guideline Group, London, United Kingdom.
24. Chopra I, Schofield C, Everett M, O'Neill A, Miller K, Wilcox M, Frère J-M, Dawson M, Czaplowski L, Urleb U, Courvalin P. 2008. Treatment of healthcare-associated infections caused by Gram-negative bacteria: a consensus statement. Lancet Infect Dis 8:133–139. [https://doi.org/10.1016/S1473-3099\(08\)70018-5](https://doi.org/10.1016/S1473-3099(08)70018-5).
25. Nielsen K, Bangsbo JM, Høiby N. 2000. Susceptibility of *Legionella* species to five antibiotics and development of resistance by exposure to erythromycin, ciprofloxacin, and rifampicin. Diagn Microbiol Infect Dis 36:43–48. [https://doi.org/10.1016/S0732-8893\(99\)00095-4](https://doi.org/10.1016/S0732-8893(99)00095-4).
26. De Giglio O, Napoli C, Lovero G, Diella G, Rutigliano S, Caggiano G, Montagna MT. 2015. Antibiotic susceptibility of *Legionella pneumophila* strains isolated from hospital water systems in southern Italy. Environ Res 142:586–590. <https://doi.org/10.1016/j.envres.2015.08.013>.
27. Descours G, Ginevra C, Jacotin N, Forey F, Chastang J, Kay E, Etienne J, Lina G, Doublet P, Jarraud S. 2017. Ribosomal mutations conferring macrolide resistance in *Legionella pneumophila*. Antimicrob Agents Chemother 61:e02188-16. <https://doi.org/10.1128/AAC.02188-16>.
28. Shadoud L, Almahmoud I, Jarraud S, Etienne J, Larrat S, Schwebel C, Timsit J-F, Schneider D, Maurin M. 2015. Hidden selection of bacterial resistance to fluoroquinolones *in vivo*: the case of *Legionella pneumophila* and humans. EBioMedicine 2:1179–1185. <https://doi.org/10.1016/j.ebiom.2015.07.018>.
29. Averbhoff B, Kirchner L, Pfefferle K, Yaman D. 2021. Natural transformation in Gram-negative bacteria thriving in extreme environments: from genes and genomes to proteins, structures and regulation. Extremophiles 25:425–436. <https://doi.org/10.1007/s00792-021-01242-z>.
30. Liu G, Thomsen LE, Olsen JE. 2022. Antimicrobial-induced horizontal transfer of antimicrobial resistance genes in bacteria: a mini-review. J Antimicrob Chemother 77:556–567. <https://doi.org/10.1093/jac/dkab450>.
31. Charpentier X, Kay E, Schneider D, Shuman HA. 2011. Antibiotics and UV radiation induce competence for natural transformation in *Legionella pneumophila*. J Bacteriol 193:1114–1121. <https://doi.org/10.1128/JB.01146-10>.
32. van Hal SJ, Wiklendt A, Espedido BR, Ginn A, Iredell JR. 2009. Immediate appearance of plasmid-mediated resistance to multiple antibiotics upon antibiotic selection: an argument for systematic resistance epidemiology. J Clin Microbiol 47:2325–2327. <https://doi.org/10.1128/JCM.02141-08>.
33. Brown-Jaque M, Calero-Cáceres W, Muniesa M. 2015. Transfer of antibiotic-resistance genes via phage-related mobile elements. Plasmid 79:1–7. <https://doi.org/10.1016/j.plasmid.2015.01.001>.
34. Massip C, Descours G, Ginevra C, Doublet P, Jarraud S, Gilbert C. 2017. Macrolide resistance in *Legionella pneumophila*: the role of LpeAB efflux pump. J Antimicrob Chemother 72:1327–1333. <https://doi.org/10.1093/jac/dkw594>.
35. Almahmoud I, Kay E, Schneider D, Maurin M. 2009. Mutational paths towards increased fluoroquinolone resistance in *Legionella pneumophila*. J Antimicrob Chemother 64:284–293. <https://doi.org/10.1093/jac/dkp173>.
36. Elliott SL, Suhrbier A, Miles JJ, Lawrence G, Pye SJ, Le TT, Rosenstengel A, Nguyen T, Allworth A, Burrows SR, Cox J, Pye D, Moss DJ, Bharadwaj M. 2008. Phase I trial of a CD8⁺ T-cell peptide epitope-based vaccine for infectious mononucleosis. J Virol 82:1448–1457. <https://doi.org/10.1128/JVI.01409-07>.
37. Mahmud A, Khan MT, Iqbal A. 2019. Identification of novel drug targets for humans and potential vaccine targets for cattle by subtractive genomic analysis of *Brucella abortus* strain 2308. Microb Pathog 137:103731. <https://doi.org/10.1016/j.micpath.2019.103731>.
38. Khan MT, Islam R, Jerin TJ, Mahmud A, Khatun S, Kobir A, Islam MN, Akter A, Mondal SI. 2021. Immunoinformatics and molecular dynamics approaches: next generation vaccine design against West Nile virus. PLoS One 16:e0253393. <https://doi.org/10.1371/journal.pone.0253393>.
39. Mondal SI, Ferdous S, Jewel NA, Akter A, Mahmud S, Islam MM, Afrin T, Karim N. 2015. Identification of potential drug targets by subtractive genome analysis of *Escherichia coli* O157: H7: an *in silico* approach. Adv Appl Bioinform Chem 8:49–63. <https://doi.org/10.2147/AABCS.88522>.
40. Barh D, Jain N, Tiwari S, Parida BP, D'Afonseca V, Li L, Ali A, Santos AR, Guimaraes LC, de Castro Soares S, Miyoshi A, Bhattacharjee A, Misra AN, Silva A, Kumar A, Azevedo V. 2011. 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 78:73–84. <https://doi.org/10.1111/j.1747-0285.2011.01118.x>.
41. Babu TM, Levine M, Fitzgerald T, Luke C, Sangster MY, Jin H, Topham D, Katz J, Treanor J, Subbarao K. 2014. Live attenuated H7N7 influenza vaccine primes for a vigorous antibody response to inactivated H7N7 influenza vaccine. Vaccine 32:6798–6804. <https://doi.org/10.1016/j.vaccine.2014.09.070>.
42. Rappuoli R. 2000. Reverse vaccinology. Curr Opin Microbiol 3:445–450. [https://doi.org/10.1016/s1369-5274\(00\)00119-3](https://doi.org/10.1016/s1369-5274(00)00119-3).
43. MacLennan CA, Mutreja A, Dougan G. 2016. New approaches for needed vaccines: bacteria, p 311–329. In *The vaccine book*. Elsevier, New York, NY.
44. Purcell AW, McCluskey J, Rossjohn J. 2007. More than one reason to rethink the use of peptides in vaccine design. Nat Rev Drug Discov 6:404–414. <https://doi.org/10.1038/nrd2224>.

45. Wong S-S, Jeevan T, Kercher L, Yoon S-W, Petkova A-M, Crumpton J-C, Franks J, Debeauchamp J, Rubrum A, Seiler P, Krauss S, Webster R, Webby RJ. 2014. A single dose of whole inactivated H7N9 influenza vaccine confers protection from severe disease but not infection in ferrets. *Vaccine* 32:4571–4577. <https://doi.org/10.1016/j.vaccine.2014.06.016>.
46. Poland GA, Ovsyannikova IG, Jacobson RM. 2009. Application of pharmacogenomics to vaccines. *Pharmacogenomics* 10:837–852. <https://doi.org/10.2217/pgs.09.25>.
47. Flower DR. 2008. *Bioinformatics for vaccinology*. John Wiley & Sons, New York, NY.
48. Xiang Z, He Y. 2013. Genome-wide prediction of vaccine targets for human herpes simplex viruses using Vaxign reverse vaccinology. *BMC Bioinform* 14:1–10. <https://doi.org/10.1186/1471-2105-14-54-52>.
49. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL. 2010. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 9:203–214. <https://doi.org/10.1038/nrd3078>.
50. Mora M, Telford JL. 2010. Genome-based approaches to vaccine development. *J Mol Med (Berl)* 88:143–147. <https://doi.org/10.1007/s00109-009-0574-9>.
51. Meunier M, Guyard-Nicodème M, Hirschaud E, Parra A, Chemaly M, Dory D. 2016. Identification of novel vaccine candidates against campylobacter through reverse vaccinology. *J Immunol Res* 2016:5715790. <https://doi.org/10.1155/2016/5715790>.
52. Barh D, Tiwari S, Jain N, Ali A, Santos AR, Misra AN, Azevedo V, Kumar A. 2011. *In silico* subtractive genomics for target identification in human bacterial pathogens. *Drug Dev Res* 72:162–177. <https://doi.org/10.1002/ddr.20413>.
53. Recanatini M, Bottegoni G, Cavalli A. 2004. *In silico* antitarget screening. *Drug Discov Today Technol* 1:209–215. <https://doi.org/10.1016/j.ddtec.2004.10.004>.
54. Bartram J, Chartier Y, Lee JV, Pond K, Surman-Lee S. 2007. Legionella and the prevention of legionellosis. World Health Organization, Geneva, Switzerland.
55. Yoo BK, Triller DM, Yong C-S, Lodise TP. 2004. Gemifloxacin: a new fluoroquinolone approved for treatment of respiratory infections. *Ann Pharmacother* 38:1226–1235. <https://doi.org/10.1345/aph.1E003>.
56. Rahimi B, Vesal A. 2017. Antimicrobial resistance properties of *Legionella pneumophila* isolated from the cases of lower respiratory tract infections. *Biomed Pharmacol J* 10:59–65. <https://doi.org/10.13005/bpj/1081>.
57. Swango K, Hymes J, Brown P, Wolf B. 2000. Amino acid homologies between human biotinidase and bacterial aliphatic amidases: putative identification of the active site of biotinidase. *Mol Genet Metab* 69: 111–115. <https://doi.org/10.1006/mgme.2000.2959>.
58. Hediger MA, Turk E, Wright EM. 1989. Homology of the human intestinal Na⁺/glucose and *Escherichia coli* Na⁺/proline cotransporters. *Proc Natl Acad Sci U S A* 86:5748–5752. <https://doi.org/10.1073/pnas.86.15.5748>.
59. Luo H, Lin Y, Liu T, Lai F-L, Zhang C-T, Gao F, Zhang R. 2021. DEG 15, an update of the Database of Essential Genes that includes built-in analysis tools. *Nucleic Acids Res* 49:D677–D686. <https://doi.org/10.1093/nar/gkaa917>.
60. Judson N, Mekalanos JJ. 2000. TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. *Nat Biotechnol* 18: 740–745. <https://doi.org/10.1038/77305>.
61. Michael CA, Dominey-Howes D, Labbate M. 2014. The antimicrobial resistance crisis: causes, consequences, and management. *Front Public Health* 2:145. <https://doi.org/10.3389/fpubh.2014.00145>.
62. Baliga P, Shekar M, Venugopal MN. 2018. Potential outer membrane protein candidates for vaccine development against the pathogen *Vibrio anguillarum*: a reverse vaccinology based identification. *Curr Microbiol* 75:368–377. <https://doi.org/10.1007/s00284-017-1390-z>.
63. Golkar Z, Bagasra O, Pace DG. 2014. Bacteriophage therapy: a potential solution for the antibiotic resistance crisis. *J Infect Dev Ctries* 8:129–136. <https://doi.org/10.3855/jidc.3573>.
64. Bartlett JG, Gilbert DN, Spellberg B. 2013. Seven ways to preserve the miracle of antibiotics. *Clin Infect Dis* 56:1445–1450. <https://doi.org/10.1093/cid/cit070>.
65. Fung M, Thornton A, Mybeck K, Wu JH-h, Hornbuckle K, Muniz E. 2001. Evaluation of the characteristics of safety withdrawal of prescription drugs from worldwide pharmaceutical markets-1960 to 1999. *Drug Information J* 35:293–317. <https://doi.org/10.1177/009286150103500134>.
66. Moses PL, Schroeder B, Alkhatib O, Ferrentino N, Suppan T, Lidofsky SD. 1999. Severe hepatotoxicity associated with bromfenac sodium. *Am J Gastroenterol* 94:1393–1396. <https://doi.org/10.1111/j.1572-0241.1999.01093.x>.
67. Andrade RJ, Lucena MI, Martin-Vivaldi R, Fernandez MC, Nogueras F, Pelaez G, Gomez-Outes A, Garcia-Escaño MD, Bellot V, Hervás J, Cárdenas F, Bermudez F, Romero M, Salmerón J. 1999. Acute liver injury associated with the use of ebrotidine, a new H2-receptor antagonist. *J Hepatol* 31:641–646. [https://doi.org/10.1016/s0168-8278\(99\)80343-2](https://doi.org/10.1016/s0168-8278(99)80343-2).
68. Lazarczyk D, Goldstein N, Gordon S. 2000. Trovafloxacin hepatotoxicity. *Am J Gastroenterol* 95:2585–2585. <https://doi.org/10.1111/j.1572-0241.2000.02959.x>.
69. Lazzaroni J-C, Dubuisson J-F, Vianney A. 2002. The Tol proteins of *Escherichia coli* and their involvement in the translocation of group A colicins. *Biochimie* 84:391–397. [https://doi.org/10.1016/S0300-9084\(02\)01419-0](https://doi.org/10.1016/S0300-9084(02)01419-0).
70. Dubuisson J-F, Vianney A, Hugouvieux-Cotte-Pattat N, Lazzaroni JC. 2005. Tol-Pal proteins are critical cell envelope components of *Erwinia chrysanthemi* affecting cell morphology and virulence. *Microbiology (Reading)* 151:3337–3347. <https://doi.org/10.1099/mic.0.28237-0>.
71. Bonsor DA, Hecht O, Vankemmelbeke M, Sharma A, Krachler AM, Housden NG, Lilly KJ, James R, Moore GR, Kleanthous C. 2009. Allosteric β -propeller signalling in TolB and its manipulation by translocating colicins. *EMBO J* 28:2846–2857. <https://doi.org/10.1038/emboj.2009.224>.
72. Clavel T, Germon P, Vianney A, Portulier R, Lazzaroni JC. 1998. TolB protein of *Escherichia coli* K-12 interacts with the outer membrane peptidoglycan-associated proteins Pal, Lpp and OmpA. *Mol Microbiol* 29: 359–367. <https://doi.org/10.1046/j.1365-2958.1998.00945.x>.
73. Nicholson TF, Watts KM, Hunstad DA. 2009. OmpA of uropathogenic *Escherichia coli* promotes postinvasion pathogenesis of cystitis. *Infect Immun* 77:5245–5251. <https://doi.org/10.1128/IAI.00670-09>.
74. Henry T, Pommier S, Journet L, Bernadac A, Gorvel JP, Llobès R. 2004. Improved methods for producing outer membrane vesicles in Gram-negative bacteria. *Res Microbiol* 155:437–446. <https://doi.org/10.1016/j.resmic.2004.04.007>.
75. Hirakawa H, Suzue K, Tomita H. 2022. Roles of the Tol/Pal system in bacterial pathogenesis and its application to antibacterial therapy. *Vaccines* 10:422. <https://doi.org/10.3390/vaccines10030422>.
76. Li Q, Li Z, Fei X, Tian Y, Zhou G, Hu Y, Wang S, Shi H. 2022. The role of TolA, TolB, and TolR in cell morphology, OMV production, and virulence of *Salmonella Choleraesuis*. *AMB Express* 12:5. <https://doi.org/10.1186/s13568-022-01347-4>.
77. Ghaffari-Nazari H, Tavakkol-Afshari J, Jaafari MR, Tahaghoghi-Hajghorbani S, Masoumi E, Jalali SA. 2015. Improving multi-epitope long peptide vaccine potency by using a strategy that enhances CD4⁺ T help in BALB/c mice. *PLoS One* 10:e0142563. <https://doi.org/10.1371/journal.pone.0142563>.
78. Aziz F, Tufail S, Shah MA, Shah MS, Habib M, Mirza O, Iqbal M, Rahman M. 2019. *In silico* epitope prediction and immunogenic analysis for penton base epitope-focused vaccine against hydropericardium syndrome in chicken. *Virus Res* 273:197750. <https://doi.org/10.1016/j.virusres.2019.197750>.
79. Yang Y, Sun W, Guo J, Zhao G, Sun S, Yu H, Guo Y, Li J, Jin X, Du L, Jiang S, Kou Z, Zhou Y. 2015. *In silico* design of a DNA-based HIV-1 multi-epitope vaccine for Chinese populations. *Hum Vaccin Immunother* 11: 795–805. <https://doi.org/10.1080/21645515.2015.1012017>.
80. Khan M, Khan S, Ali A, Akbar H, Sayaf AM, Khan A, Wei D-Q. 2019. Immunoinformatics approaches to explore *Helicobacter pylori* proteome (virulence factors) to design B and T cell multi-epitope subunit vaccine. *Sci Rep* 9:13. <https://doi.org/10.1038/s41598-019-49354-z>.
81. Ishii KJ, Akira S. 2008. Innate immunity, p 39–51. *In Clinical immunology*. Elsevier, New York, NY.
82. Shim HK, Kim JY, Kim MJ, Sim HS, Park DW, Sohn JW, Kim MJ. 2009. Legionella lipoprotein activates Toll-like receptor 2 and induces cytokine production and expression of costimulatory molecules in peritoneal macrophages. *Exp Mol Med* 41:687–694. <https://doi.org/10.3858/emmm.2009.41.10.075>.
83. Brieland J, Freeman P, Kunkel R, Chrisp C, Hurley M, Fantone J, Engleberg C. 1994. Replicative *Legionella pneumophila* lung infection in intratracheally inoculated A/J mice: a murine model of human Legionnaires' disease. *Am J Pathol* 145:1537–1546.
84. Baskerville A, Fitzgeorge RB, Broster M, Hambleton P, Dennis PJ. 1981. Experimental transmission of Legionnaires' disease by exposure to aerosols of *Legionella pneumophila*. *Lancet* 2:1389–1390. [https://doi.org/10.1016/s0140-6736\(81\)92803-8](https://doi.org/10.1016/s0140-6736(81)92803-8).
85. Wright EK, Goodart SA, Grownen JD, Hadinoto V, Endrizzi MG, Long EM, Sadigh K, Abney AL, Bernstein-Hanley I, Dietrich WF. 2003. Naip5 affects host susceptibility to the intracellular pathogen *Legionella pneumophila*. *Curr Biol* 13:27–36. [https://doi.org/10.1016/S0960-9822\(02\)01359-3](https://doi.org/10.1016/S0960-9822(02)01359-3).

86. Padilla-Carlin DJ, McMurray DN, Hickey AJ. 2008. The guinea pig as a model of infectious diseases. *Comp Med* 58:324–340.
87. Komura T, Yasui C, Miyamoto H, Nishikawa Y. 2010. *Caenorhabditis elegans* as an alternative model host for *Legionella pneumophila*, and protective effects of *Bifidobacterium infantis*. *Appl Environ Microbiol* 76: 4105–4108. <https://doi.org/10.1128/AEM.03021-09>.
88. Kubori T, Shinzawa N, Kanuka H, Nagai H. 2010. *Legionella* metaeffector exploits host proteasome to temporally regulate cognate effector. *PLoS Pathog* 6:e1001216. <https://doi.org/10.1371/journal.ppat.1001216>.
89. Harding CR, Schroeder GN, Collins JW, Frankel G. 2013. Use of *Galleria mellonella* as a model organism to study *Legionella pneumophila* infection. *J Vis Exp* 2013:e50964. <https://doi.org/10.3791/50964>.
90. Abu Kwak Y. 1996. The phagosome containing *Legionella pneumophila* within the protozoan *Hartmannella vermiformis* is surrounded by the rough endoplasmic reticulum. *Appl Environ Microbiol* 62:2022–2028. <https://doi.org/10.1128/aem.62.6.2022-2028.1996>.
91. Chien M, Morozova I, Shi S, Sheng H, Chen J, Gomez SM, Asamani G, Hill K, Nuara J, Feder M, Rineer J, Greenberg JJ, Steshenko V, Park SH, Zhao B, Teplitskaya E, Edwards JR, Pampou S, Georghiou A, Chou I-C, Iannuccilli W, Ulz ME, Kim DH, Geringer-Sameth A, Goldsberry C, Morozov P, Fischer SG, Segal G, Qu X, Rzhetsky A, Zhang P, Cayanis E, De Jong PJ, Ju J, Kalachikov S, Shuman HA, Russo JJ. 2004. The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* 305:1966–1968. <https://doi.org/10.1126/science.1099776>.
92. Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26: 680–682. <https://doi.org/10.1093/bioinformatics/btq003>.
93. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, Berry A, Bhai J, Bignell A, Billis K, Boddu S, Brooks L, Charkhchi M, Cummins C, Da Rin Fioretto L, Davidson C, Dodiya K, Donaldson S, et al. 2022. Ensembl 2022. *Nucleic Acids Res* 50:D988–D995. <https://doi.org/10.1093/nar/gkab1049>.
94. Shanmugham B, Pan A. 2013. Identification and characterization of potential therapeutic candidates in emerging human pathogen *Mycobacterium abscessus*: a novel hierarchical *in silico* approach. *PLoS One* 8: e59126. <https://doi.org/10.1371/journal.pone.0059126>.
95. Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30. <https://doi.org/10.1093/nar/28.1.27>.
96. Qiu Y, Yu-Qing Q. 2013. KEGG pathway database pp. 1068–1069. *In* Encyclopedia of System Biology. Springer New York, NY, USA.
97. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35:W182–W185. <https://doi.org/10.1093/nar/gkm321>.
98. Li YH, Xu JY, Tao L, Li XF, Li S, Zeng X, Chen SY, Zhang P, Qin C, Zhang C, Chen Z, Zhu F, Chen YZ. 2016. SVM-Prot 2016: a web-server for machine learning prediction of protein functional families from sequence irrespective of similarity. *PLoS One* 11:e0155290. <https://doi.org/10.1371/journal.pone.0155290>.
99. Damte D, Suh J-W, Lee S-J, Yohannes SB, Hossain MA, Park S-C. 2013. Putative drug and vaccine target protein identification using comparative genomic analysis of KEGG annotated metabolic pathways of *Mycoplasma hyopneumoniae*. *Genomics* 102:47–56. <https://doi.org/10.1016/j.ygeno.2013.04.011>.
100. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FSL. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26: 1608–1615. <https://doi.org/10.1093/bioinformatics/btq249>.
101. Yu CS, Lin CJ, Hwang JK. 2004. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on *n*-peptide compositions. *Protein Sci* 13:1402–1406. <https://doi.org/10.1110/ps.03479604>.
102. King BR, Vural S, Pandey S, Barteau A, Guda C. 2012. ngLOC: software and web server for predicting protein subcellular localization in prokaryotes and eukaryotes. *BMC Res Notes* 5:351. <https://doi.org/10.1186/1756-0500-5-351>.
103. Bhasin M, Garg A, Raghava GPS. 2005. PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics* 21:2522–2524. <https://doi.org/10.1093/bioinformatics/bti309>.
104. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 46:D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>.
105. Doytchinova IA, Flower DR. 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform* 8:4. <https://doi.org/10.1186/1471-2105-8-4>.
106. Fatoba AJ, Okpeku M, Adeleke MA. 2021. Subtractive genomics approach for identification of novel therapeutic drug targets in *Mycoplasma genitalium*. *Pathogens* 10:921. <https://doi.org/10.3390/pathogens10080921>.
107. Sears CL. 2005. A dynamic partnership: celebrating our gut flora. *Anaerobe* 11:247–251. <https://doi.org/10.1016/j.anaerobe.2005.05.001>.
108. Wadood A, Jamal A, Riaz M, Khan A, Uddin R, Jelani M, Azam SS. 2018. Subtractive genome analysis for *in silico* identification and characterization of novel drug targets in *Streptococcus pneumoniae* strain JJA. *Microb Pathog* 115:194–198. <https://doi.org/10.1016/j.micpath.2017.12.063>.
109. Guarner F, Malagelada J-R. 2003. Gut flora in health and disease. *Lancet* 361:512–519. [https://doi.org/10.1016/S0140-6736\(03\)12489-0](https://doi.org/10.1016/S0140-6736(03)12489-0).
110. Savage DC. 1977. Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol* 31:107–133. <https://doi.org/10.1146/annurev.mi.31.100177.000543>.
111. Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H, Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, Peavy H, Pontzer C, Portnoy M, Sayre MH, Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M, NIH HMP Working Group. 2009. The NIH human microbiome project. *Genome Res* 19:2317–2323. <https://doi.org/10.1101/gr.096651.109>.
112. Human Microbiome Project Consortium. 2012. A framework for human microbiome research. *Nature* 486:215–221. <https://doi.org/10.1038/nature11209>.
113. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C. 2021. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 49:D605–D612. <https://doi.org/10.1093/nar/gkaa1074>.
114. Kuntal BK, Aparoy P, Reddanna P. 2010. EasyModeller: a graphical interface to MODELLER. *BMC Res Notes* 3:226. <https://doi.org/10.1186/1756-0500-3-226>.
115. Xu D, Zhang Y. 2011. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys J* 101:2525–2534. <https://doi.org/10.1016/j.bpj.2011.10.024>.
116. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291. <https://doi.org/10.1107/S0021889892009944>.
117. Colovos C, Yeates TO. 1993. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 2:1511–1519. <https://doi.org/10.1002/pro.5560020916>.
118. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. 2005. Protein identification and analysis tools on the ExPASy server, p 571–607. *In* The proteomics protocols handbook. Humana Press, Totowa, NJ. <https://doi.org/10.1385/1-59259-890-0571>.
119. Buus S, Lauemøller S, Worning P, Kesmir C, Frimurer T, Corbet S, Fomsgaard A, Hilden J, Holm A, Brunak S. 2003. Sensitive quantitative predictions of peptide-MHC binding by a “Query by Committee” artificial neural network approach. *Tissue Antigens* 62:378–384. <https://doi.org/10.1034/j.1399-0039.2003.00112.x>.
120. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>.
121. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. 2006. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinform* 7:153. <https://doi.org/10.1186/1471-2105-7-153>.
122. Dimitrov I, Bangov I, Flower DR, Doytchinova I. 2014. AllerTOP v.2: a server for *in silico* prediction of allergens. *J Mol Model* 20:2278–2276. <https://doi.org/10.1007/s00894-014-2278-5>.
123. Dimitrov I, Naneva L, Doytchinova I, Bangov I. 2014. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* 30:846–851. <https://doi.org/10.1093/bioinformatics/btt619>.
124. Chrysostomou C, Seker H. 2014. Prediction of protein allergenicity based on a signal-processing bioinformatics approach, p 808–811. *In* Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, Chicago, IL. <https://doi.org/10.1109/EMBC.2014.6943714>.
125. Fiers MW, Kleter GA, Nijland H, Peijnenburg AA, Nap JP, Van Ham RC. 2004. Allermatch™, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinform* 5:133–136. <https://doi.org/10.1186/1471-2105-5-133>.

126. Sharma N, Naorem LD, Jain S, Raghava GPS. 2022. ToxinPred2: an improved method for predicting toxicity of proteins. *Brief Bioinform* 2022:bbac174. <https://doi.org/10.1093/bib/bbac174>.
127. Bui HH, Sidney J, Li W, Füsseder N, Sette A. 2007. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinform* 8:361. <https://doi.org/10.1186/1471-2105-8-361>.
128. Kolaskar AS, Tongaonkar PC. 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 276: 172–174. [https://doi.org/10.1016/0014-5793\(90\)80535-q](https://doi.org/10.1016/0014-5793(90)80535-q).
129. Karplus P, Schulz G. 1985. Prediction of chain flexibility in proteins. *Naturwissenschaften* 72:212–213. <https://doi.org/10.1007/BF01195768>.
130. Chou P, Fasman G. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol* 47:45–148. <https://doi.org/10.1002/9780470122921.ch2>.
131. Emini EA, Hughes JV, Perlow D, Boger J. 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 55: 836–839. <https://doi.org/10.1128/JVI.55.3.836-839.1985>.
132. Parker J, Guo D, Hodges R. 1986. 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* 25:5425–5432. <https://doi.org/10.1021/bi00367a013>.
133. Jespersen MC, Peters B, Nielsen M, Marcotilli P. 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 45:W24–W29. <https://doi.org/10.1093/nar/gkx346>.
134. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, Lundegaard C, Sette A, Lund O, Bourne PE, Nielsen M, Peters B. 2012. Immune epitope database analysis resource. *Nucleic Acids Res* 40: W525–W530. <https://doi.org/10.1093/nar/gks438>.
135. Rana A, Akhter Y. 2016. A multi-subunit based, thermodynamically stable model vaccine using combined immunoinformatics and protein structure based approach. *Immunobiology* 221:544–557. <https://doi.org/10.1016/j.imbio.2015.12.004>.
136. Pandey RK, Sundar S, Prajapati VK. 2016. Differential expression of miRNA regulates T cell differentiation and plasticity during visceral leishmaniasis infection. *Front Microbiol* 7:206. <https://doi.org/10.3389/fmicb.2016.00206>.
137. Hajjigharamani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatbadi SS, Ghasemi Y. 2017. Immunoinformatics analysis and *in silico* designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. *Infect Genet Evol* 48:83–94. <https://doi.org/10.1016/j.meegid.2016.12.010>.
138. Sharma N, Patiyal S, Dhall A, Pande A, Arora C, Raghava GPS. 2021. AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes. *Brief Bioinform* 22.
139. McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* 16:404–405. <https://doi.org/10.1093/bioinformatics/16.4.404>.
140. Petersen B, Lundegaard C, Petersen TN. 2010. NetTurnP: neural network prediction of beta-turns by use of evolutionary information and predicted protein sequence features. *PLoS One* 5:e15079. <https://doi.org/10.1371/journal.pone.0015079>.
141. Peng J, Xu J. 2011. RaptorX: exploiting structure information for protein alignment by statistical inference. *Proteins* 79:161–171. <https://doi.org/10.1002/prot.23175>.
142. Källberg M, Margaryan G, Wang S, Ma J, Xu J. 2014. RaptorX server: a resource for template-based protein structure modeling, p 17–27. *In* Protein structure prediction. Springer, New York, NY.
143. Zhang J, Liang Y, Zhang Y. 2011. Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. *Structure* 19:1784–1795. <https://doi.org/10.1016/j.str.2011.09.022>.
144. Lovell SC, Davis IW, Arendall WB, III, De Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. 2003. Structure validation by $C\alpha$ geometry: ϕ , ψ , and $C\beta$ deviation. *Proteins* 50:437–450. <https://doi.org/10.1002/prot.10286>.
145. Craig DB, Dombkowski AA. 2013. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinform* 14:346–347. <https://doi.org/10.1186/1471-2105-14-346>.
146. Solanki V, Tiwari V. 2018. Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against *Acinetobacter baumannii*. *Sci Rep* 8:1–19. <https://doi.org/10.1038/s41598-018-26689-7>.
147. Ahmadishoar S, Kariminik A. 2017. Toll-like receptor 2 and its roles in immune responses against *Legionella pneumophila*. *Life Sci* 188:158–162. <https://doi.org/10.1016/j.lfs.2017.09.009>.
148. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. 2004. ClusPro: a fully automated algorithm for protein–protein docking. *Nucleic Acids Res* 32: W96–W99. <https://doi.org/10.1093/nar/gkh354>.
149. Yan Y, Zhang D, Zhou P, Li B, Huang SY. 2017. HDock: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res* 45:W365–W373. <https://doi.org/10.1093/nar/gkx407>.
150. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. 2005. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 33:W363–W367. <https://doi.org/10.1093/nar/gki481>.
151. Wüthrich K, Wagner G, Richarz R, Braun W. 1980. Correlations between internal mobility and stability of globular proteins. *Biophys J* 32: 549–560. [https://doi.org/10.1016/S0006-3495\(80\)84989-7](https://doi.org/10.1016/S0006-3495(80)84989-7).
152. Van Aalten DM, De Groot BL, Findlay JB, Berendsen HJ, Amadei A. 1997. A comparison of techniques for calculating protein essential dynamics. *J Comput Chem* 18:169–181. [https://doi.org/10.1002/\(SICI\)1096-987X\(19970130\)18:2%3C169::AID-JCC3%3E3.0.CO;2-T](https://doi.org/10.1002/(SICI)1096-987X(19970130)18:2%3C169::AID-JCC3%3E3.0.CO;2-T).
153. Tama F, Brooks CL, III. 2006. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annu Rev Biophys Biomol Struct* 35: 115–133. <https://doi.org/10.1146/annurev.biophys.35.040405.102010>.
154. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29:845–854. <https://doi.org/10.1093/bioinformatics/btt055>.
155. Huang J, MacKerell JA. 2013. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* 34: 2135–2145. <https://doi.org/10.1002/jcc.23354>.
156. Berweger CD, van Gunsteren WF, Müller-Plathe F. 1995. Force field parametrization by weak coupling: re-engineering SPC water. *Chem Physics Lett* 232:429–436. [https://doi.org/10.1016/0009-2614\(94\)01391-8](https://doi.org/10.1016/0009-2614(94)01391-8).
157. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. 2018. PDBsum: structural summaries of PDB entries. *Protein Sci* 27:129–134. <https://doi.org/10.1002/pro.3289>.
158. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D. 2005. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* 33:W526–W531. <https://doi.org/10.1093/nar/gki376>.