scientific reports

Check for updates

OPEN

Core-genome guided novel therapeutic targets identification and chimeric vaccine designing against *Rickettsia rickettsii*

Fizza Arshad[1,8], Asifa Sarfraz[1,8], Muhammad Shehroz[2], Umar Nishan[3], Asia Perveen[4], Riaz Ullah[5], Mohamed A. Ibrahim[6] & Mohibullah Shah[1,7]✉

Rocky Mountain Spotted Fever, caused by the gram-negative intracellular bacteria *Rickettsia rickettsii*, is a serious tick-borne infection with a fatality rate of 20–30%, if not treated. Since it is the most serious rickettsial disease in North America, modified prevention and treatment strategies are of critical importance. In order to find new therapeutic targets and create multiepitope vaccines, this study integrated subtractive proteomics with reverse vaccinology. The core genome of *R. rickettsii* was investigated, resulting in the identification of seven essential, human non-homologous proteins as potential drug targets, as well as four antigenic, non-allergenic proteins suitable for vaccine development. Using conserved antigenic peptides, two chimeric vaccine constructs were developed and assessed using molecular docking, molecular dynamics simulations, principal component analysis, MM-GBSA binding free energy, and dynamic cross-correlation matrix studies. The high immunogenic potential was indicated by the vaccine designs' robust and consistent interactions with human immunological receptors. Their capacity to trigger strong humoral and cellular immunological responses was further demonstrated by in silico immune simulations. The persistent interactions of vaccine V1 and V2 with human immunological receptor demonstrated that these might have high immunogenic potential. Moreover, the identified drug targets were annotated for essential biological processes, which shed light on their therapeutic potential. The vaccine constructs were cloned and expressed in suitable systems. This study displays a comprehensive strategy for managing Rocky Mountain Spotted Fever via rational vaccine development. Further experimental research is needed to confirm the immunogenicity of the vaccines and the druggability of identified targets, establishing the path toward effective RMSF management.

Keywords Immunoinformatics, reverse vaccinology, drug pocket, binding energy, Brazilian spotted fever

The genus *Rickettsia* is one of the most diverse and common genera among the several intracellular gram-negative bacteria that make up the Rickettsiaceae family[1]. One of the most common and serious rickettsial disorders in the United States (U.S.) is Rocky Mountain Spotted Fever (RMSpF), which is a member of the spotted fever group rickettsioses (SFGR) family of diseases. The highly pathogenic intracellular bacteria *Rickettsia rickettsii (R. rickettsii)* is the cause of this disorder[2]. This pathogen is equally harmful to humans and other species. Ticks, the organism's vectors, and reservoirs, play a vital role in its natural ecosystem. Human infection is caused by the bite of an infected tick *Dermacentor andersoni* and *Dermacentor variabilis*, which are the carriers for *R. rickettsii*[3]. SFGR is becoming more recognized as a serious global public health risk due to its emerging and re-emerging phenomena. The majority of cases of *R. rickettsii* were discovered either in or during the second half of the 20th century, with the first case being recorded in 1906 [4]. Since its identification in different regions in 2003, RMSF has remained a growing concern to public health in the southwestern part of the United States[5]. Symptoms of

[1]Department of Biochemistry, Bahauddin Zakariya University, Multan 66000, Punjab, Pakistan. [2]Department of Bioinformatics, Kohsar University Murree, Murree 47150, Pakistan. [3]Department of Chemistry, Kohat University of Science & Technology, Kohat, Pakistan. [4]School of Natural Sciences, Faculty of Science and Engineering, Macquarie University, Macquarie Park, NSW 2109, Australia. [5]Department of Pharmacognosy, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia. [6]Department of Pharmaceutics, College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia. [7]Department of Animal Science, Federal University of Ceara, Fortaleza, Brazil. [8]Fizza Arshad and Asifa Sarfraz contributed equally to this work. ✉email: mohib@bzu.edu.pk; mohibusb@gmail.com

SFGR typically include fever, chills, malaise, and myalgia. If treatment is not started immediately, the disease can worsen and cause complications like shock, coagulopathy, respiratory distress, abdominal pain, rash, and altered mental status, all of which can appear in as little as five days. Although rashes appear early higher death rates are associated with delayed rash onset because of the delayed treatment[6].

Over the years, multiple outbreaks of RMSpF, have been reported. First documented in 1937 in Tobia, then, the illness reappeared in the vicinity of Tobia in 2003 and among military personnel in Antioquia later in 2005. On the northern coast of Colombia, in the municipality of Los Córdobas, in the Córdoba department, there was a new epidemic in 2007. Over the Americas, RMSF is a zoonosis that is widely spread[7]. Despite interdisciplinary control efforts, a rickettsioses outbreak that started in Nuevo Leon, Mexico, in 2022 is still ongoing[8]. Several other infectious and non-infectious disorders frequently overlap with non-specific symptoms like fever, headaches, and malaise, making it difficult to diagnose rickettsial diseases. Furthermore, a low level of bacteremia is associated with the early acute stage of sickness, which makes it more difficult to identify the causative agent[9].

The Centers for Disease Control and Prevention (CDC) and The American Academy of Pediatrics (AAP) recommended doxycycline as the first line of treatment for patients of all ages who have suspected rickettsial diseases[10]. Non-specific presentation of rickettsial infections might delay identification and treatment with doxycycline leading to serious consequences and death[11]. Furthermore, the development of a vaccine to prevent *R. rickettsii* poses a significant social impact because there are currently no commercial methods for diagnosing the acute phase of the disease, which is primarily found in vulnerable communities far away from specialized health services and causes patients to become critically ill and eventually die[12]. Meanwhile, there is no vaccine available to protect against RMSF in dogs or humans. It is complicated to understand the protective host response and the *R. rickettsii* antigens that stimulate protective immunity, so developing vaccines against RMSF is challenging[13]. This enabled it to design novel therapeutic and vaccine targets that would prove useful in the near future. The objective of this study is to discover new targets for drugs and vaccines against this bacterium to offer substitute therapeutic strategies. The strategy used in this study is noteworthy, which makes use of bioinformatics tools to reduce labor and concentrate on developing vaccine candidates' predictions.

Materials and methods

In this study, we utilized state-of-the-art computational tools (Fig. 1) to rigorously analyze the core proteins of *R. rickettsii* genomes in search of novel therapeutic targets.

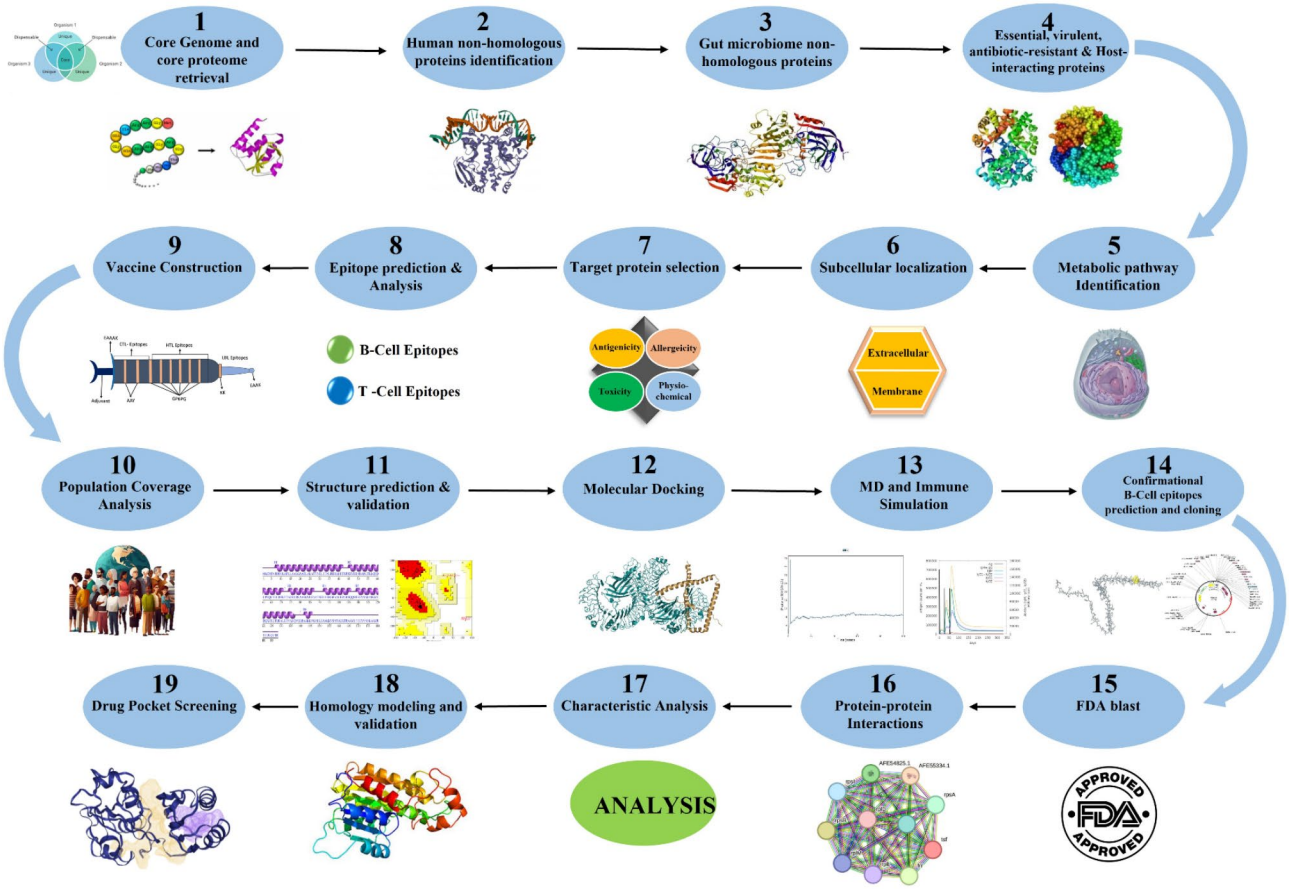

Fig. 1. Schematic illustration of the methodology used to create the MEVCs against *R. rickettsii*.

Data retrieval

The complete genome sequences of all the available 15 strains of *R. rickettsii* were retrieved from the National Center for Biotechnology Information (NCBI) database. These genome sequences were then submitted to EDGAR tool version 2 for the core genome formulation[14].

Determination of human non-homologous proteins

The therapeutic targets should not be identical to the human proteins to prevent auto-immunity[15]. The complete human proteome was retrieved from the UniProt database[16] And standalone BLASTp analysis of core proteome was accomplished against human proteome to determine human host non-homologous proteins[17]. The cutoff values used for BLASTp analysis include query coverage < 35%, percentage identity < 35%, bitscore < 100, E-value and > 1e-20[18,19]. Furthermore, the human digestive system is populated by many symbiotic bacteria, all of which have numerous beneficial effects on human health. Developing a vaccine or drug that targets proteins similar to those in the gut microbiota carries the risk of adversely affecting commensal microorganisms living in the gastrointestinal tract, which could have negative effects on human physiological health[20,21]. BLASTp analysis of the human non-homologous proteins was performed to determine gut-non-homologous proteins by applying similar cutoff values.

Selection of essential proteins and virulent factors

Essential genes encode proteins that are essential to basic cellular functions like transcription, translation, metabolism, and DNA replication. Essential genes are particularly interesting in the context of drug development or vaccine target identification because blocking or altering their function may cause important cellular processes to be disrupted, which could ultimately lead to the removal of pathogens or the treatment of diseases. The DEG (Database of Essential Genes) contains genes that are essential for the survival of the pathogen. The DEG library comprises experimentally validated vital genes from bacteria, archaea, and eukaryotes. Moreover, the pathogenicity of bacteria is influenced by virulence factors that strengthen their capacity to invade, infect, and elude host defenses. The VFDB (Virulence Factor Database) provides a comprehensive knowledge base and a versatile platform for the evaluation of bacterial virulence factors. These virulent factors play a crucial role in the development and spread of pathogenic diseases[21,22]. To identify essential & virulent proteins of *R. rickettsii*, the human non-homologous proteins were screened against the DEG and VFDB databases with thresholds i.e. qcovs > 35%, percentage identity > 35%, bitscore > 100, E-value 1e-20 [21,23].

Selection of resistant and host interacting proteins

Bacterial infections are becoming resistant to currently available antibiotics. Therefore, the Antibiotic-Resistant Gene Annotation (ARG-ANNOT) database was used to identify proteins that help in the resistance of *R. rickettsii* to antibiotics by BLASTp analysis[24]. Moreover, the Host-Pathogen Protein-Protein Interaction database (HPPPI) played a crucial role in the selection of host-interacting proteins of the *R. rickettsii*[25]. The bacterial proteins were subjected to PPI for the identification of the proteins interacting with the host counterparts.

Metabolic pathways analysis

The KAAS (KEGG Automatic Annotation Server) server was used to assign KAAS IDs to the pathogen essential, virulent, resistant, and host-interacting proteins. The KEGG (Kyoto Encyclopedia of Genes and Genomes) server was used to analyze the metabolic pathways of *R. rickettsii* unique proteins[26]. The pathways of *R. rickettsii* which are unique to humans were identified and the proteins involved in those unique pathways were used for further analysis.

Identification of subcellular localization of target proteins

Identifying the subcellular location of proteins is essential in recognizing their roles within cells. Based on the localization of proteins within the cell, they are classified as either drug or vaccine targets. In this study, PSORTb v3.0, CELLOv2.5, and BUSCA servers were used for the subcellular localization of the KEGG-dependent and independent proteins. The cytoplasmic proteins were classified as potential drug targets while the extracellular and membrane proteins were considered vaccine targets and used for further downstream analysis[27,28].

Reverse vaccinology

Identifying promising vaccine candidates involves a multifaceted approach that incorporates computational methods for the analysis of several properties of target proteins. To identify antigenic and non-allergenic proteins, VaxiJen[29] and AllerTOP[30] web servers were used, respectively. Protein sequences containing transmembrane helices were elucidated via the TMHMM server[31], thus rendering it to identify proteins with topological values of 0 or 1. Important physicochemical properties such as molecular weight, amino acid count, theoretical isoelectric point (Pi), instability index, aliphatic index, half-life, and the grand average of hydropathicity (GRAVY) value were determined by ProtParam Expasy server[32]. The promising vaccine candidates were selected based on their ability to exhibit antigenicity and non-allergenicity, as well as their negative GRAVY value and topological value of 0 or 1[33].

Prediction of HTL epitopes

The most important cells in adaptive immunity, helper T-cells (HTL) influence B-cells to release antibodies, macrophages to phagocytose pathogens, and cytotoxic T-cells (CTL) to destroy specific parasitized cells that are required for cell-mediated immunization against *Rickettsia rickettsii*. Therefore, to stimulate the immune response, the inclusion of HTL epitopes is required. The selected vaccine candidates were assessed to predict HTL epitopes with a length of 15 mer, using the IEDB MHC-II binding epitope prediction tool[34,35]. During

the prediction of epitopes, the HLA full reference allele set was selected from IEDB. The top 10 epitopes with the lowest rank were selected and were evaluated for their antigenicity, allergenicity, solubility, virulence, and toxicity analysis. Furthermore, HTL epitopes generate various cytokines, including interleukin-10 (IL-10), interferon-gamma (IFN-γ), and interleukin-4 (IL-4), which stimulate immune cells including cytotoxic T-cells and macrophages. These additional properties were determined by IL10-pred, IFN-epitope, and IL4-pred servers.

Prediction of CTL epitopes

The majority of cytotoxic T-cells display T-cell receptors (TCRs), which recognize a specific antigen[36]. The consensus method was utilized to identify CTL epitopes using the IEDB MHC-I binding epitope prediction tool and employing a length of 9–10 mer[34]. The epitopes used for vaccine construction must be antigenic, non-allergenic, immunogenic, non-toxic, virulent, and water-soluble. The top 10 epitopes with the lowest rank were evaluated for these properties using the VaxiJen 2.0, AllerTOP, IEDB, Toxinpred, Virulentpred, and Innovagen servers, respectively.

Prediction of B-cell epitope

A multi-epitope vaccine should activate humoral immunity to produce long-lasting humoral protection. B-cell epitopes-induced humoral immunity can eradicate infections by generating antibodies against antigens found in the human body. The effectiveness of vaccines is due to memory B-cells specific to antigens, which stimulate the formation of antibodies (Abs). Additionally, these contribute to the removal of pathogens by phagocytic or complement-mediated mechanisms[37,38]. The consensus method was utilized to identify B-cell epitopes using the BepiPred Linear Epitope prediction server. The epitopes with a length of 10 to 40 residues were evaluated for their antigenicity, allergenicity, toxicity, virulence, and solubility.

Multi-epitope vaccine construction

The vaccine was formulated by the use of the top prioritized epitopes[39]. Additionally, various linkers were added, the bi-lysine (KK) linker for B-cell epitopes, Gly-Pro-Gly-Pro-Gly (GPGPG) for HTL epitopes, and Ala-Ala-Tyr (AAY) for CTL epitopes to connect with the prioritized epitopes[40]. Additionally, a rigid linker i.e., EAAAK was included at the C and N-terminals of the vaccine constructs as well as at the point of adjuvant and epitopes attachment. In addition, adjuvants that may stimulate the immune system, such as HBHA and HBHA conserved adjuvants, were used to enhance the immunological responses and promote innate immunity. After the vaccine's construction, their physicochemical properties such as molecular weight, number of amino acids, theoretical pi, instability index, allergenicity, and antigenicity were analyzed. The water solubility, antigenicity, and allergenicity of the designed vaccine constructs were determined by the SolPro[41], VaxiJen, AntigenPro, and AllerTOP servers, respectively.

Population coverage analysis

A subset of epitopes from $CD8^+$ and $CD4^+$ T-cells along with their corresponding HLA alleles were submitted to the IEDB Population Coverage tool. This tool predicts the percentage of HLA alleles and epitopes that are compatible with the selected populations. The global population coverage was also determined to determine the global efficacy of the vaccines[42].

Structure analysis

The secondary structures of the designed vaccine constructs were formulated using the SOPMA[43] and PsiPred[44] Web servers. These servers provided evidence regarding the percentage of random coils, extended strands, beta-turn, and alpha-helix in the vaccine construct. Moreover, the 3D structures of the vaccines were formed using the SWISS-MODEL server[45]. The resulting 3D structures were refined using the GalaxyRefine web server[46]. The refined structures were then validated by the Ramachandran plot and the ERRAT quality factor value[47].

Molecular docking analysis

Molecular docking was performed to ascertain the binding affinities and molecular interactions of the proposed vaccines with the human toll-like immune receptor, TLR4 (PDB-ID: 3FXI). The receptor and vaccines were prepared for docking. The prepared structures were submitted to the Cluspro2.0 server for molecular docking[48]. The binding affinities of the vaccine-receptor complexes were determined and the Pymol software displayed the interactions among the docked complexes. For additional research, the vaccines with the highest binding energy and maximum number of H-bond interactions with TLR4 were selected.

Normal mode analysis

The mobility of macromolecules at large scales and their stability were investigated by employing the normal mode analysis mobility. This is done to validate the stability of the vaccine-receptor docked complexes. iMODS server computes the NMA of the dihedral coordinates of the Cα atoms in addition to a specified combined motion of a large macromolecule. In addition, iMODS calculates the eigenvalue, evaluates structural deformability, B-factor, variance, and covariance[49].

Molecular dynamic simulation

Molecular dynamics (MD) simulation is typically used to evaluate the stability of a complex containing prospective therapeutic compounds and the target protein[50–52]. Molecular dynamics simulations were conducted for 100 ns by using the Desmond software package by Schrödinger LLC[53]. The protein-ligand complexes were processed beforehand including optimization and minimization, with either the Protein Preparation Wizard or Maestro

software, before conducting the simulations. The System Builder tool was used to facilitate system construction. To simulate realistic environmental conditions, we employed the TIP3P solvent model within an orthorhombic simulation box. The simulations were run using the OPLS_2005 force field[50], and the models were neutralized by adding counter ions when necessary. To mimic physiological circumstances, we added 0.15 M salt. To simulate physiological circumstances, 0.15 M salt solution (NaCl) was added. We employed the NPT ensemble throughout the simulation to ensure that moles (N), pressure (P), and temperature (T) remained constant at 300 K and 1 atm, respectively. Pre-simulation relaxing procedures were carried out on the models. To assess simulation stability, we calculated the root mean square deviation (RMSD), and root mean square fluctuations (RMSF) to identify the changes in the structure and stability of the proteins throughout the simulations for the vaccine-receptor complexes.

Principal component analysis

Using the bio3d R package, the PCA of each complex was obtained from the MD trajectories[54]. This statistical technique minimizes and determines substantial fluctuations in protein residues using covariance matrix analysis of alpha-carbon (Cα) atoms. The eigenvectors that have the highest eigenvalues are considered principal components. This approach takes into account three key components. It provides information on the dynamics of the protein by computing its principal components[55].

MM-GBSA energy calculation

The molecular mechanics generalized Born surface area (MM-GBSA) module of prime was used to determine the binding free energy (Gbind) of the docked complex during MD simulations of V1 and V2 complexed with TLR4. Using the OPLS 2005 force field, VSGB solvent model, and rotamer search techniques, the binding free energy was estimated. The MD trajectory frames were chosen at intervals of 10 ns after the MD run.

DCCM analysis

The dynamic cross-correlation matrix (DCCM) analysis was conducted to investigate the correlated motions of residues within the V1-TLR4 and V2-TLR4 docked complex over a 100 ns molecular dynamics (MD) simulation. The trajectory data from the MD simulation was processed to calculate the cross-correlation coefficients between the fluctuations of each pair of residues. The resulting DCCM was visualized as a heatmap, with colors ranging from −1.0 (indicating anti-correlated motions) to 1.0 (indicating positively correlated motions). This analysis allowed for the identification of regions within the complex that exhibited significant correlated or anti-correlated movements, providing insights into the dynamic behavior and potential functional interactions of the vaccine-receptor docked complexes.

Conformational B-cell epitopes

The Conformational B-cell epitopes in the 3D structure of the prioritized vaccine construct were obtained by using the ElliPro webserver[56] which assists in antibody prediction against protein antigens based on their sequence. As an ellipsoid, it employs the most advanced approaches for protein structural analysis.

Immune simulation

To evaluate the immunological profile and effectiveness of the prioritized vaccines, immune simulation analysis was performed using the C-ImmSim server. The immune simulations were performed at 1000 steps, and 3 injections were employed at different time intervals (01, 84, and 168 h)[57].

In silico expression analysis

The codons in the vaccine construct were optimized following the expression system of *E. coli* strain K12 using the JCAT tool[58].The vaccine sequences were back-translated, with the adaptation mediated by the values of the Codon Adaptation Index (CAI) and the percentage of GC content. During optimization, de-selected regions included bacterial ribosome binding sites, rho-independent transcription terminators, and several restriction enzyme cleavage sites. After that, the SnapGene tool was used to conduct in-silico cloning in the pET-28a (+) vector of the *E. coli* host to ensure the expression of the vaccine construct[38].

Novel drug targets identification

The shortlisted cytoplasmic proteins underwent BLASTp analysis against FDA-approved drug targets with a bit score of at least 100 and an E-value of less than 0.005 to identify the novel proteins[59,60]. To predict the novel targets, the proteins showing similarity with the FDA-approved drug targets were removed from the study. The remaining proteins were analyzed for their protein-protein interactions by the STRING database[61] with an average node degree greater than 5.0 ($K \geq 5$), transmembrane helix prediction by the TMHMM server[31], and physicochemical properties (molecular weight prediction, instability index, theoretical pi, and aliphatic index) prediction by the Protparam tool. The low molecular weight, stable, interacting, and proteins with less than 1 transmembrane helices were shortlisted as potential novel drug targets[62].

3D structure prediction and catalytic pocket screening

The Swiss Model server was utilized to anticipate the 3D structures of the prioritized targets. The designed structures were validated by analyzing the ERRAT score and Ramachandran plot using the SAVES server[63]. Following the acquisition of 3D structures, the DoGSiteScorer, an automated pocket detection technology to determine if protein cavities are druggable, was utilized to determine the drug-binding pockets and their druggability score. A highly druggable protein cavity is defined as having a score closer to 1.

Results

Core proteome retrieval

The proteins of 17 available complete genomes of *Rickettsia rickettsii* were retrieved from the NCBI database and subjected to the EDGAR server for core genome development, which resulted in 1214 core proteins that were used for subtractive proteomics and the reverse vaccinology analysis[14].

Determination of human and gut microbiome non-homologous proteins

For the identification of novel therapeutic targets, 1214 core proteins were subjected to BLASTp analysis against the *Homo sapiens* proteome. This resulted in the selection of 1121 proteins because they were non-homologous to human proteins. Since, the gut microbiota is crucial for digesting and other vital body functions, proteins linked to gut flora were excluded from the analysis. As a result, 984 pathogen-specific proteins were obtained which were used for further analysis.

Essential, virulent, antibiotic-resistant, and host-interacting proteins

The BLASTp analysis resulted in the 133 essential proteins required for *R. rickettsii* survival and maintenance, crucial in determining novel vaccine targets. Virulence proteins are essential for regulating the host defense mechanism through colonization, invasion, adhesion, and pathogenicity. According to the VFDB database, 16 proteins were linked to *R. rickettsii* virulence.

ARG-ANNOT database analysis resulted in 3 antibiotic-resistant proteins. The HPPPI database was used to identify host-interacting proteins. Three proteins were found to be involved in the interaction of bacteria with the host proteins. Following the integration of essential, virulent, antibiotic-resistant, and host-interacting proteins, 145 potential therapeutic targets were determined which were used for further analysis. These proteins might be the key targets for preventing *R. rickettsii* infection.

Determination of unique metabolic pathways

The metabolic pathways and their proteins were identified by the KEGG server. The total number of metabolic pathways of *R. rickettsii* and humans was found to be 77 and 356, respectively. Of these, 59 metabolic pathways were common in both humans and *R. rickettsia*, while 18 pathways were present only in *R. rickettsii* (Table 1). After the determination of unique pathways, the proteins from the previous analysis were subjected to the KAAS server for assignment of KO (KEGG orthology) identifiers to each protein. Out of 145 proteins, 133 proteins were identified as KEGG-dependent while the remaining 12 proteins were found to be KEGG-independent. The 133 kegg-dependent proteins were mapped to the pathogen pathways, which resulted in 47 pathogen pathways, of which 10 pathways were found to be unique to the pathogen. The 39 proteins (Table S1) involved in these 10 unique metabolic pathways were processed for further analysis.

Identification of subcellular localization of target proteins

Predicting subcellular localization allows for the identification of appropriate and efficient targets, which is necessary for an extensive understanding of protein function. Cytoplasmic proteins are preferred as effective drug targets while the membrane proteins are considered the best candidates for vaccines. The result of the subcellular localization indicated that out of 51 proteins, 31 were cytoplasmic while 20 were membranous proteins (Table 2; Fig. 2), that were subjected to drug and vaccine targets identification, respectively.

Sr. No.	Entry	Description
1	rri00261	Monobactam biosynthesis
2	rri00300	Lysine biosynthesis
3	rri00460	Cyanoamino acid metabolism
4	rri00521	Streptomycin biosynthesis
5	rri00540	Lipopolysaccharide biosynthesis
6	rri00541	O-Antigen nucleotide sugar biosynthesis
7	rri00550	Peptidoglycan biosynthesis
8	rri00680	Methane metabolism
9	rri00946	Degradation of flavonoids
10	rri01110	Biosynthesis of secondary metabolites
11	rri01120	Microbial metabolism in diverse environments
12	rri01220	Degradation of aromatic compounds
13	rri01501	Beta-Lactam resistance
14	rri01502	Vancomycin resistance
15	rri01503	Cationic antimicrobial peptide (CAMP) resistance
16	rri02020	Two-component system
17	rri02024	Quorum sensing
18	rri03070	Bacterial secretion system

Table 1. Pathways of *Rickettsia rickettsii* unique to Host (Humans).

Sr. No.	Accession numbers	Protein name	Localization
1	QBX69_03605	MULTISPECIES: 30 S ribosomal protein S15	Cytoplasmic
2	QBX69_03545	MULTISPECIES: Ribonucleotide-diphosphate reductase subunit beta	Cytoplasmic
3	QBX69_00550	MULTISPECIES: dCTP deaminase	Cytoplasmic
4	QBX69_06495	MULTISPECIES: SH3 domain-containing protein	Cytoplasmic
5	QBX69_06055	MULTISPECIES: 50 S ribosomal protein L21	Cytoplasmic
6	QBX69_05270	MULTISPECIES: 50 S ribosomal protein L22	Cytoplasmic
7	QBX69_00185	MULTISPECIES: CarD family transcriptional regulator	Cytoplasmic
8	QBX69_04835	MULTISPECIES: Chromosomal replication initiator protein DnaA	Membrane
9	QBX69_05105	MULTISPECIES: Nucleotide exchange factor GrpE	Cytoplasmic
10	QBX69_00180	DsbA family protein	Cytoplasmic
11	QBX69_00365	30 S ribosomal protein S6	Cytoplasmic
12	QBX69_00570	Cell cycle transcriptional regulator TrcR	Cytoplasmic
13	QBX69_00635	Translation elongation factor Ts	Cytoplasmic
14	QBX69_00715	ABC transporter permease	Membrane
15	QBX69_00760	Ribosome biogenesis GTP-binding protein YihA/YsxC	Membrane
16	QBX69_00830	Protein translocase subunit SecF	Membrane
17	QBX69_00885	FtsH protease activity modulator HflK	Cytoplasmic
18	QBX69_00895	Do family serine endopeptidase	Membrane
19	QBX69_00900	Rhodanese domain-containing protein	Cytoplasmic
20	QBX69_00970	50 S ribosomal protein L10	Cytoplasmic
21	QBX69_01210	6-pyruvoyl tetrahydropterin synthase family protein	Cytoplasmic
22	QBX69_01295	TrbC/VirB2 family protein	Membrane
23	QBX69_02345	Endolytic transglycosylase MltG	Membrane
24	QBX69_02715	ATP-binding cassette domain-containing protein	Cytoplasmic
25	QBX69_02905	Holliday junction branch migration protein RuvA	Cytoplasmic
26	QBX69_02935	D-alanyl-D-alanine carboxypeptidase family protein	Membrane
27	QBX69_02940	Septal ring lytic transglycosylase RlpA family protein	Membrane
28	QBX69_02975	TlpA disulfide reductase family protein	Membrane
29	QBX69_03020	Cytochrome c oxidase subunit I	Membrane
30	QBX69_03235	MULTISPECIES: response regulator transcription factor	Membrane
31	QBX69_04650	Holo-ACP synthase	Cytoplasmic
32	QBX69_05830	Ribonuclease HI	Cytoplasmic
33	QBX69_06170	ABC-F family ATP-binding cassette domain-containing protein	Cytoplasmic
34	QBX69_06185	ATP-binding protein	Cytoplasmic
35	QBX69_06355	MULTISPECIES: NADH-quinone oxidoreductase subunit NuoK	Membrane
36	QBX69_06660	Outer membrane protein	Membrane
37	QBX69_06675	MULTISPECIES: heme ABC transporter permease	Membrane
38	QBX69_06825	Bifunctional peptide chain release factor N(5)-glutamine methyltransferase	Cytoplasmic

Table 2. Shortlisted cytoplasmic and membrane proteins.

Reverse vaccinology

Vaccine candidate prioritization

The membrane proteins prioritized in the previous step analyzed through the TMHMM server to find out the topology value of the proteins. Several characteristics (i.e., instability index, aliphatic index, PI, GRAVY, and molecular weight) of all the vaccine target membrane proteins were analyzed by the ProtParam tool. The antigenicity, allergenicity, solubility, and virulence were also determined (Table 3). All these analyses resulted in four proteins including QBX69_00760, QBX69_00895, QBX69_02940, and QBX69_03235 as the best vaccine candidates for further utilization in vaccine designing due to their best properties such as they were antigenic, non-allergenic, stable, virulent, water-soluble, and have a low molecular weight (Table 3).

Prediction of lead epitopes

The IEDB-recommended NetMHCpan EL 4.0 web server was used to predict CTL and HTL epitopes. For this purpose, the entire collection of HLA reference alleles from IEDB was used. Many MHC-I and MHC-II binding epitopes for the proteins QBX69_00760, QBX69_00895, QBX69_02940, and QBX69 _03235 were obtained. The top 10 epitopes of MHC-I and MHC-II with the lowest rank from each protein were evaluated for toxicity, antigenicity, virulence, allergenicity, and water solubility. Furthermore, the IEDB Class-I immunogenicity prediction method was employed to assess the immunogenicity of the MHC-I-binding epitopes. After careful consideration, one epitope each from QBX69_00760, QBX69_00895, and QBX69 _03235 was chosen for vaccine

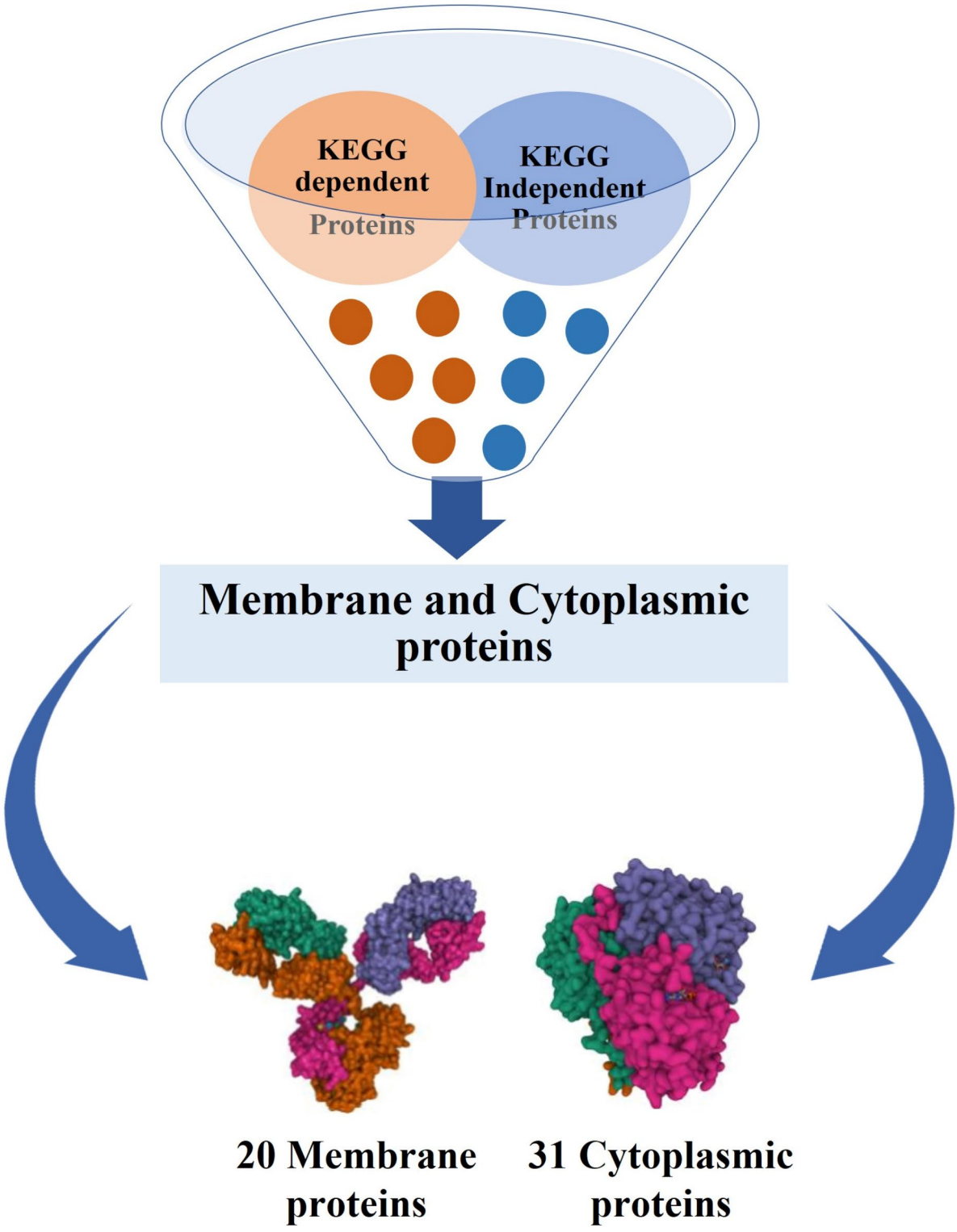

Fig. 2. Graphical representation of the final localization of the kegg-dependent and independent proteins.

development because of their positive immunogenicity, non-toxicity, non-allergenicity, antigenicity, solubility, and virulent qualities. After analysis, three epitopes each from QBX69_00760 and QBX69_00895 and one epitope from QBX69_02940, and QBX69_03235 were chosen for vaccine development because their virulent, non-toxic, non-allergenic, and antigenic potentials, as well as inducing the cytokines production (Table S2, S3).

Moreover, B-cell epitopes were determined using the BePipred linear epitope prediction server from the IEDB repository. Epitopes that met the recommended length criteria (10 to 40) were analyzed for virulence, antigenicity, toxicity, allergenicity, and water solubility. Seven linear B-cell epitopes (non-toxic, non-allergenic,

Protein ID	No. of amino acids	Topology	Molecular Weight	Theoretical Pi	Aliphatic Index	GRAVY	Antigen	Allergen	Instability index	Water solubility
QBX69_04835	463	0	52942.68	6.85	96.83	−0.243	No	No	Unstable	Good
QBX69_00715	259	i13-35o50-72i151-173o200-222i235-257o	27670.75	9.34	124.56	0.917	Yes	No	Stable	Poor
QBX69_00760	212	0	23874.63	9.89	94.29	−0.284	Yes	No	Stable	Good
QBX69_00830	308	i20-42o138-160i165-187o192-214i245-267o272-294i	35075.55	9.43	121.49	0.515	Yes	No	Stable	Poor
QBX69_00895	508	0	55606.23	4.66	106.65	−0.232	Yes	No	Stable	Good
QBX69_01295	119	i21-43o53-75i82-104o	12803.25	7.76	139.16	0.789	Yes	No	Stable	Poor
QBX69_02345	338	i12-34o	38001.77	9.68	109.88	−0.013	No	No	Stable	Good
QBX69_02935	306	0	33603.64	10.29	85.2	−0.2	Yes	Yes	Stable	Good
QBX69_02940	320	0	36020.81	9.9	83.22	−0.585	Yes	No	Stable	Good
QBX69_02975	214	i9-31o	24976.09	8.9	102.48	−0.155	No	No	Stable	Poor
QBX69_03020	532	i31-53o75-97i118-140o165-187i200-222o254-276i283-305o320-342i355-377o392-414i427-449o469-491i	58909.37	6.56	101.37	0.691	Yes	No	Stable	Poor
QBX69_03235	226	0	25612.51	8.89	104.87	−0.244	Yes	No	Stable	Good
QBX69_06355	104	o5-27i34-56o66-88i	11780.08	5.46	139.71	0.939	Yes	No	Unstable	Poor
QBX69_06660	246	0	26502.24	9.64	77.28	−0.211	Yes	Yes	Stable	Good
QBX69_06675	230	i20-42o57-79i86-108o123-145i157-179o201-223i	26008.63	9.71	141.52	1.007	Yes	No	Stable	Poor
QBX69_07145	145	o10-32i53-75o80-102i123-141o	16875.43	9.45	125.03	0.769	Yes	No	Stable	Poor
QBX69_02485	604	0	69066.45	8.14	97.83	−0.225	No	No	Unstable	Good
QBX69_04745	555	i134-156o176-198i205-227o232-254i360-382o397-419i431-448o453-475i488-507o522-544i	61373.1	9.3	124.47	0.686	Yes	No	Stable	Poor
QBX69_05680	415	i23-45o274-296i316-338o381-403i	46027.08	9.45	117.01	0.404	Yes	No	Stable	Poor
QBX69_05555	282	i19-36o46-68i89-111o116-133i138-160o165-187i212-229o239	32020.88	9.18	143.97	0.894	Yes	No	Unstable	Poor

Table 3. Analysis of vaccine targets obtained by subtractive proteomics and potential vaccine candidates selection.

water-soluble virulent, and antigenic) were ultimately chosen from these proteins and added to the vaccine construct (Table S4).

Vaccine construction

The lead epitopes were joined using suitable linkers (AAY, GPGPG, KK) to form peptide vaccine constructs. Furthermore, adjuvants were added to the N-terminal of the designed construct to improve its immunogenicity and form two vaccine constructs (V1& V2). The EAAAK linker was also used to provide rigidity to the ends. The effects of various adjuvant combinations were used to analyze the allergenicity and antigenicity affected by adjuvants (Fig. 3). The addition of linkers did not alter the conformation of the designed constructs. After the formation of vaccine constructs, they were analyzed for their antigenicity, allergenicity, and water solubility which indicated that both the constructs were antigenic, non-allergenic, and water-soluble. Moreover, the physicochemical properties (i.e., instability index, GRAVY value, molecular weight, amino acid length, aliphatic index, and theoretical pi) of vaccine constructs were analyzed which indicated that the designed vaccines are stable, have negative GRAVY value, and low molecular weight. The topology of both constructs was found to be 0 (Table 4). Since both vaccine constructs fulfilled the standard criteria, they were selected for further investigation.

Structural analysis of vaccine constructs

The residues involved in the synthesis of random coils, β-sheets, α-helices, and extended strands were determined by the secondary structure of the vaccines. The vaccine's overall structural stability was influenced by these specific secondary structures. The percentages of α-helix (50.26), extended strands (10.77), and random coils (35.21) for V1 were computed. Comparably, the percentages of these secondary structures for V2 were in the order of 50.69, 11.28, and 34.20, respectively (Fig. 4).

Tertiary structure prediction provides a critical analysis of the stability and function of proteins. The SWISS-MODEL was used for the prediction of the tertiary structure generating Ramachandran plot to validate the improved 3D model. A high-quality 3D structure of our vaccines was demonstrated by the Ramachandran plot analysis, which revealed 94.0% and 89.9% residues in the most preferred regions of V1 and V2, respectively (Fig. S1). Furthermore, the Pymol v2.5.5 was used to visualize the tertiary structures to improve depiction (Fig. 5).

Population coverage analysis

The MHC-I and II epitopes selected for vaccine construction along with their respective binding HLA-alleles were subjected to population coverage analysis. The global coverage of the selected T-cell epitopes was 99.74%.

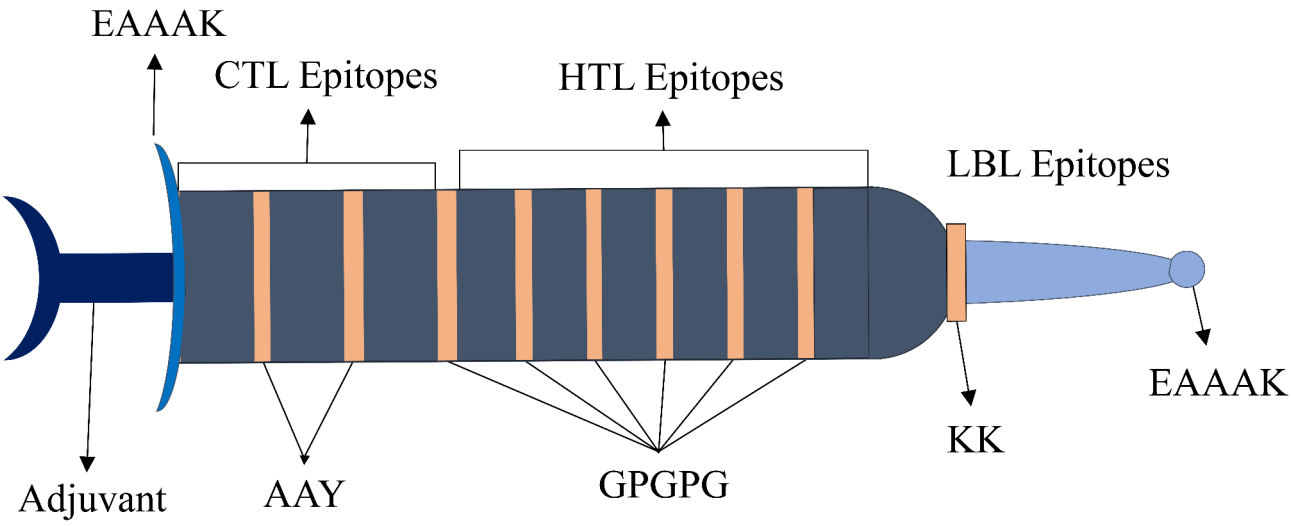

Fig. 3. Construction of MEVC joined by various linkers (AAY, GPGPG, KK) and adjuvants (HBHA and HBHA-conserved).

Parameters	V1		V2	
	Measurement	Indication	Measurement	Indication
Solpro	0.925778	Soluble	0.930743	Soluble
Antigen pro	0.94323	Antigenic	0.945135	Antigenic
VaxiJen v2.0	Antigen	Antigenic	Antigen	Antigenic
Allergenicity	Non-Allergen	Non-allergen	Non-Allergen	Non-allergen
Instability index	33.64	Stable	36.32	Stable
Molecular weight	64364.87	Appropriate	63246.64	Appropriate
GRAVY value	−0.887	Hydrophilic	-0.876	Hydrophilic
Amino acid length	585	Appropriate	576	Appropriate
Aliphatic Index	71.47	Thermostable	73.07	Thermostable
Theoretical Pi	8.14	Basic	6.99	Basic
Topology	0	Appropriate	0	Appropriate

Table 4. Physicochemical properties, antigenicity and allergenicity analysis of the two designed vaccine constructs V1 and V2.

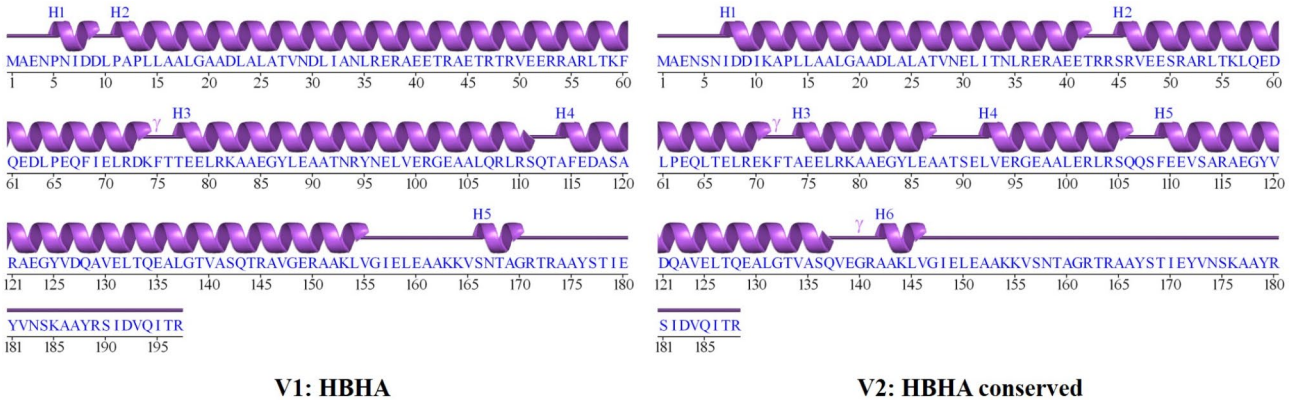

Fig. 4. Secondary structures of the designed vaccine constructs formulated by the PDBsum tool.

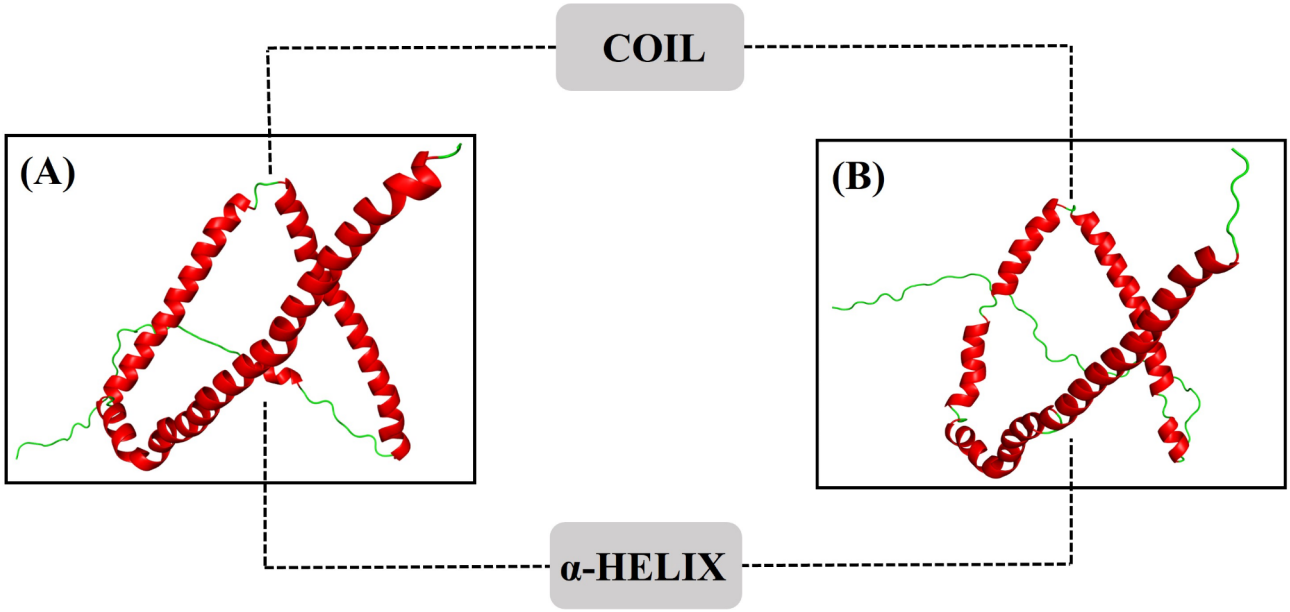

Fig. 5. Structural representation of the designed vaccine constructs indicating alpha helix (red) and random coil (green) by using the Pymol Software.

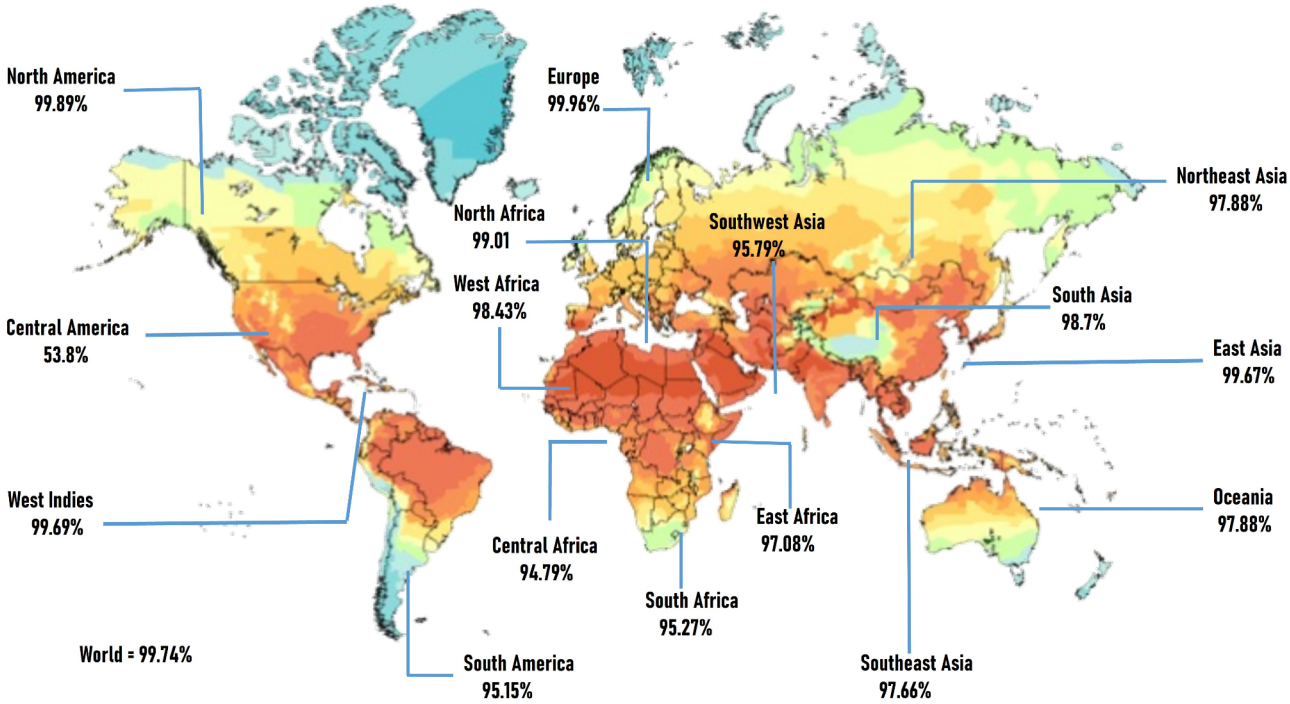

Fig. 6. Population coverage graph of T-cell epitopes used in the construction of MEVC determined by the IEDB Population Coverage Tool. The epitopes selected for vaccine construction showed a global allele coverage of 99.74%.

It was found that Central America had the lowest range of population coverage (53.8%), while Europe had the highest range of population coverage (99.96%) (Fig. 6).

Molecular docking and normal mode analysis

The ClusPro 2.0 server was used to dock the designed vaccines (V1 and V2) with TLR4 receptors to determine their binding affinities and possible interactions among them. The docked complexes of V1 and V2 with human TLR4 showed binding affinities of -1296.1 and −1294.6, respectively. Furthermore, the docked complexes and

the interactions between the vaccines and the receptors were visualized using the PyMOL software (Figs. 7 and 8). Normal mode analysis (NMA) was performed to analyze the stability of docked complexes and their large-scale mobility (Fig. S2). According to the normal mode analysis mobility, upon binding, the vaccine construct and the TLR-receptors were notably directed. The deformability of the complex was linked to the distortion of individual residues, which served as the chain's hinges. Similar to RMS, the B-factor values estimated by NMA were significantly lower than their PDB B-factor. The predicted eigenvalue for the vaccine constructs was found to have an inverse connection with the variance of the protein-protein docking complex. The estimated eigenvalues for vaccine constructs V1 and V2 with TLR-4 receptor were 1.777387e-07 and 9.132412e-07, respectively. The graphical depiction of the covariance matrix shows the correlated, uncorrelated, and anti-correlated pairs of amino acid residues, respectively, with white, red, and blue color variations (Fig. S2). In the elastic network model, springs of atomic contact are displayed as grey dots; stiffer springs are represented by darker grey, and vice versa.

Molecular dynamics simulations

The molecular dynamics simulations were carried out to check the stability of the docked complexes of the vaccine and the receptor. As a result, the RMSD and RMSF graphs were obtained (Fig. 9) which were analyzed to check the MD trajectories. The predicted alterations in their conformation from the starting structure for the whole simulation duration were represented. Similar to the current investigation, minor deviations from the RMSD curve for vaccines V1 and V2 in complex with the TLR4 receptor signify the stability of the docked complex and vice versa. For the vaccines V1-TLR4 complex and V2-TLR4 complex, the calculated RMSD is 10 Å ± 1 Å and 13 Å ± 1. In the case of the V1-TLR4 complex, only an initial deviation was observed between 0 and 4 ns. After that, the complex remained stable indicating the stability of the docked complex. Whereas, for the V2-TLR4 complex, initial deviations were obtained till 16 ns and after that no significant increase or decrease was observed and the complex remained stable ensuring the stability of the complex. Significant variations in RMSF values were found in both complexes because of the presence of intrinsically flexible regions. This is substantially linked to the loop regions attached at the N-terminus, containing inserted epitopes and linker sequences alongside the adjuvants. The local alterations along the protein chain can be explained using RMSF. The regions of the protein that fluctuate the most during the simulations are represented by peaks in the RMSF graphs. The residual flexibility study was performed to obtain a better understanding of the stability of the complexes that were generated. Figure 8 demonstrates that the flexibility values of the V1-TLR4 and V2-TLR4 are lower, which confirms the stability and efficacy of our analysis.

Principal component analysis

V1-TLR4 complex The PCA analysis of the MD simulation data for the V1 vaccine with the TLR4 receptor provides valuable insights into the structural and dynamic behavior of the complex. The results indicate that the V1-TLR4 complex exhibits multiple stable conformations, as evidenced by the distinct clusters observed in the PCA plots (Fig. 10). The first component (PC1) captures the most significant variation, suggesting that major conformational changes or dominant interactions are driving the primary structural differences in the complex. The significant proportion of variance explained by the first three principal components (67.58%) underscores the effectiveness of these components in capturing the essential features of the V1-TLR4 complex. The sharp decline in the scree plot after the first component further supports the importance of PC1 in explaining the dataset's variability. The well-defined clusters in the PCA plots suggest that the V1-TLR4 complex adopts multiple stable conformations, which is indicative of its structural stability during the simulation. The presence of these stable states, along with the substantial variance captured by the primary principal components, implies that the

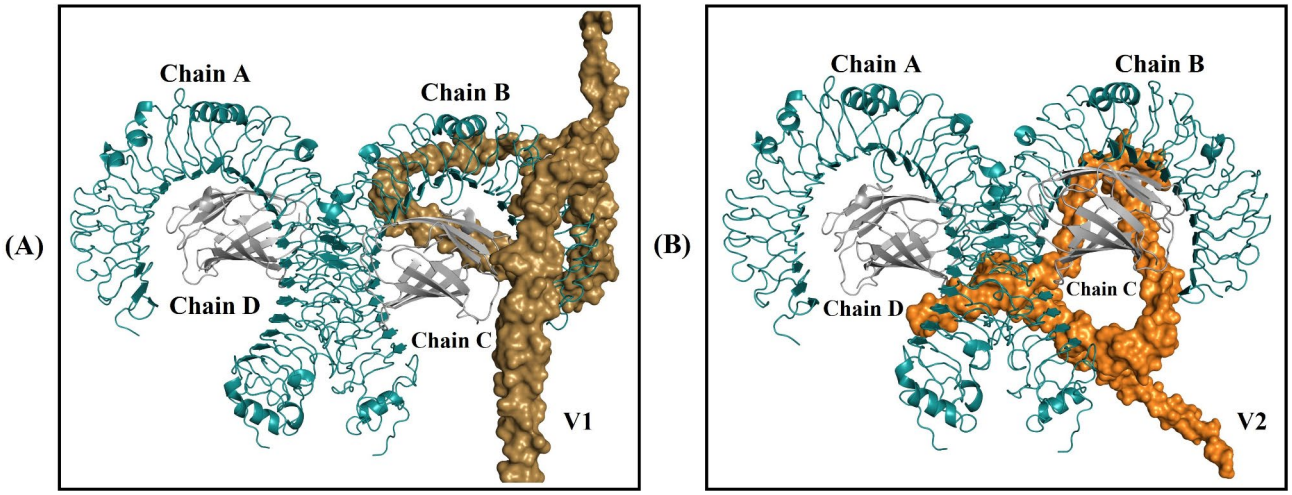

Fig. 7. Structural representation of the vaccine-receptor docked complex. The docking was performed by the Cluspro 2.0 server and visualized by the Pymol Software.

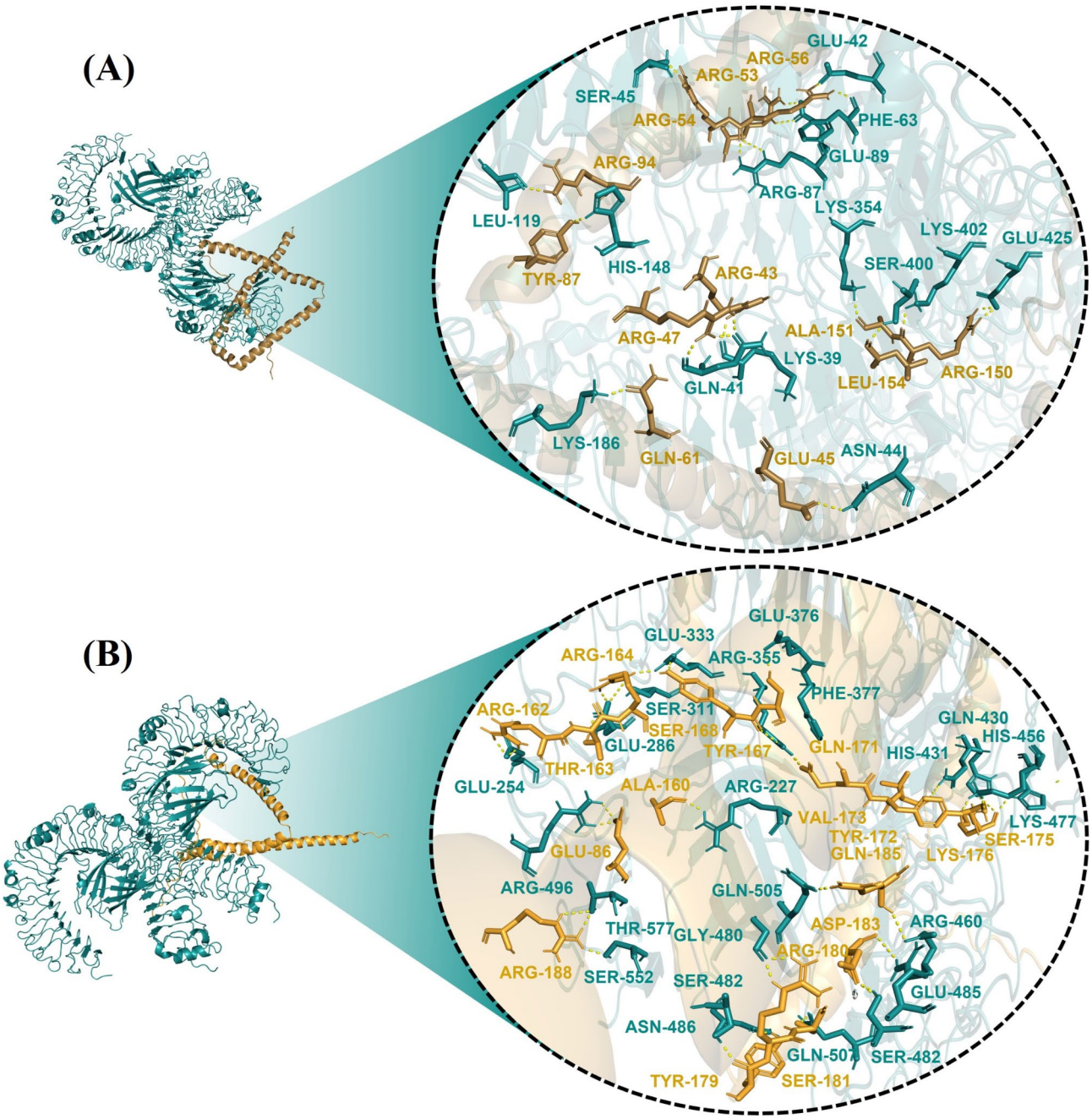

Fig. 8. Molecular Interactions between (**A**) V1 (brown) and TLR4 (zinc) (**B**) V2 (orange) and TLR4 (zinc) visualized by the Pymol Software.

complex is not only stable but also exhibits a degree of conformational flexibility that may be essential for its biological function.

V2-TLR4 complex The PCA analysis of the V2 vaccine with the TLR4 receptor provides a clear understanding of the structural and dynamic behavior of the complex (Fig. 11). The first component (PC1) captures a significant 52.2% of the total variance, indicating that major conformational changes or dominant interactions are the primary sources of structural differences in the V2-TLR4 complex. The second (PC2) and third (PC3) components capture additional dimensions of variability, explaining 14.79% and 6.05% of the total variance, respectively. These components reflect secondary and tertiary structural features or interactions, providing further details about the stability and dynamics of the complex. The distinct clusters observed in the PCA plots indicate that the V2-TLR4 complex adopts multiple stable conformations during the simulation. The separation along the PC1 axis suggests that these primary conformational states are significantly different, while the variations along PC2 and PC3 provide insights into the finer structural adjustments within these states.

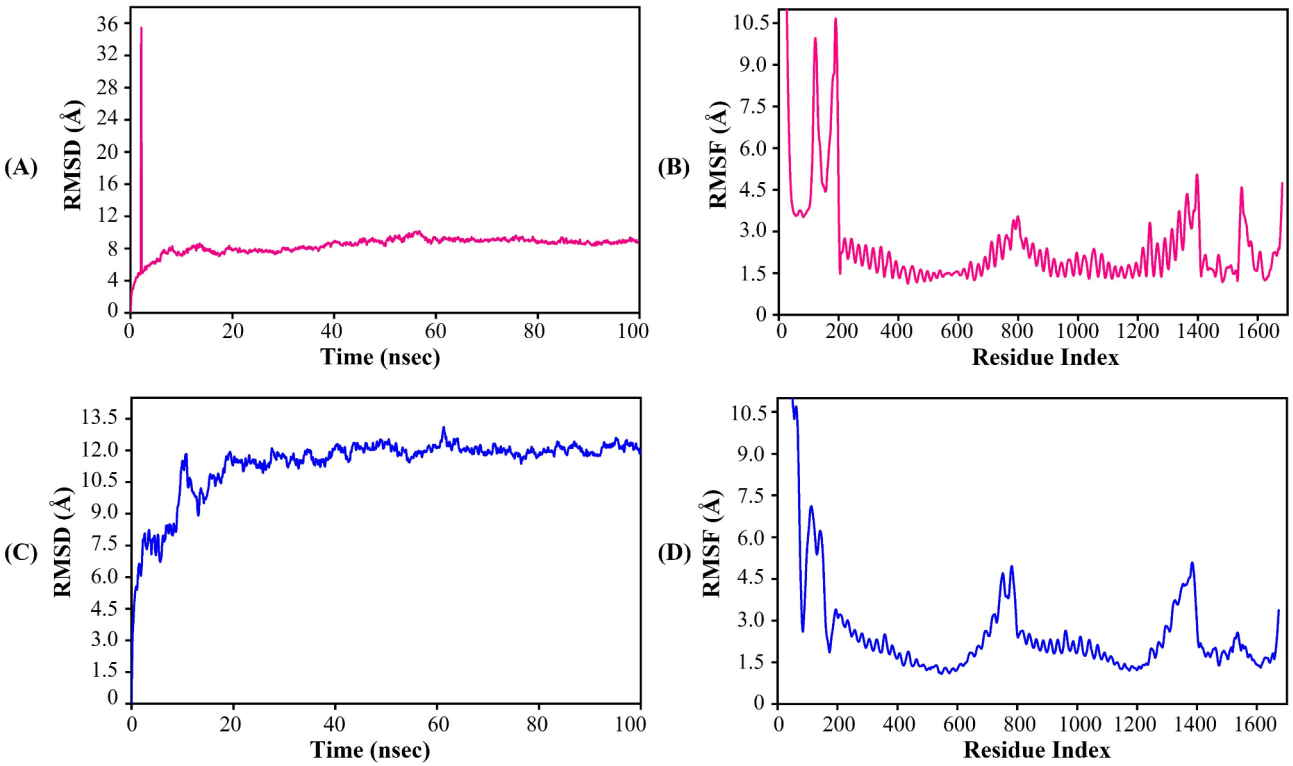

Fig. 9. Molecular Dynamic Simulation was performed by Schrodinger and resulted in RMSD and RMSF graphs for the designed vaccines and TLR4 complexes. V1 graphs (Pink) are indicated in (**A**,**B**) while the V2 (Blue) are in (**C**,**D**).

Based on the PCA analysis, it can be concluded that the V2 and TLR4 complex is stable. The multiple stable conformations observed during the simulation, along with the substantial variance explained by the primary principal components, indicate that the complex maintains its structural integrity while allowing for necessary conformational adjustments. This stability is crucial for the effective interaction of the V2-TLR4 complex, supporting its potential efficacy as an immunotherapy agent. The higher variance explained by PC1 in the V2-TLR4 complex compared to V1-TLR4 suggests that V2 may exhibit more pronounced conformational changes, which could be indicative of its interaction dynamics. However, both V1 and V2 demonstrate stability in their respective complexes with TLR4, which is essential for their intended biological functions.

Comparison of V1-TLR4 and V2-TLR4 complex

Based on the PCA analysis, the V2-TLR4 complex appears to be more stable than the V1-TLR4 complex. This conclusion is drawn from the higher variance explained by the principal components and the more pronounced separation of clusters in the PCA plots for the V2-TLR4. The dominance of PC1 in capturing the variance further supports the stability of the V2-TLR4 complex, indicating that it has more defined and stable conformational states during the simulation. Thus, the V2-TLR4 complex demonstrates greater stability and conformational clarity compared to the V1-TLR4 complex, making it a potentially more effective vaccine candidate.

Binding energy investigation

V1-TLR4 complex The analysis of the binding interactions between V1 and TLR4 reveals that the system's overall binding energy (ΔG_bind) is -240.37 kcal/mol (Table 5), indicating a highly favorable interaction between the ligand V1 and the receptor TLR4. The lipophilic contribution to the binding energy (ΔG_bind_Lipo) is -131.91 kcal/mol, highlighting the significant role of hydrophobic interactions in stabilizing the V1-TLR4 complex. Van der Waals interactions (ΔG_bind_vdW) also contribute substantially, with a value of -220.62 kcal/mol, indicating strong attractive forces at close range. The Coulombic contribution (ΔG_bind_Coulomb) is 115.28 kcal/mol, suggesting repulsive electrostatic interactions due to like-charged groups within the binding interface. Hydrogen bonding interactions (ΔG_bind_Hbond) contribute −28.13 kcal/mol, indicating the formation of hydrogen bonds that enhance the stability of the complex. Lastly, the packing interactions (ΔG_bind_Packing) contribute a minor −3.57 kcal/mol, suggesting that V1 fits well into the TLR4 binding pocket with minimal energy required for conformational adjustments (Table 5). These findings collectively underscored the various forces contributing to the binding affinity of V1 with TLR4.

V2-TLR4 complex The molecular dynamics simulation results of V2 binding with TLR4 reveal several key interaction energies (Table 5). The overall binding free energy (dG_bind) is calculated to be -65.7268 kcal/

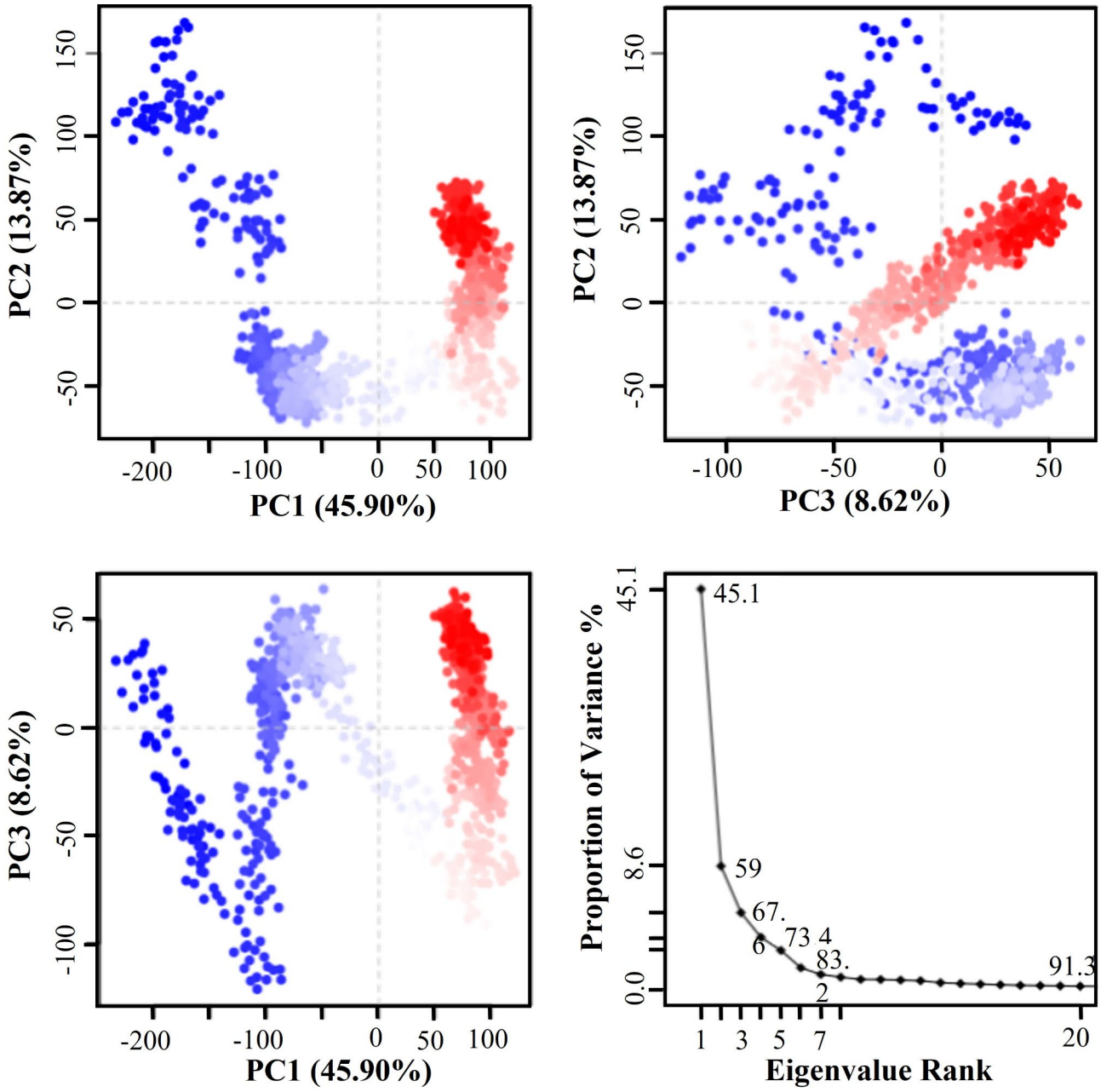

Fig. 10. PCA analysis of the V1-TLR4 complex performed by the R script.

mol. The contribution from lipophilic interactions (dG_bindLipo) is -17.0467 kcal/mol, indicating significant hydrophobic interactions between V2 and TLR4. The van der Waals interactions (dG_bindvdW) contribute −42.1463 kcal/mol to the binding energy, highlighting the importance of close contact and steric fit. Coulombic interactions (dG_bindCoulomb) are also substantial, contributing −49.3653 kcal/mol, which underscores the role of electrostatic forces in stabilizing the complex. Hydrogen bonding interactions (dG_bindH_bond) contribute −5.07422 kcal/mol, indicating specific directional bonding that further stabilizes the binding. Finally, packing interactions (dG_bindPacking) contribute −3.5666 kcal/mol, reflecting the structural complementarity and efficient packing of V2 within the TLR4 binding site. Overall, these interactions collectively facilitated a stable and energetically favorable binding of V2 with TLR4.

Comparison of V1-TLR4 and V2-TLR4 complex When comparing the binding interactions of V1 and V2 with TLR4, it is evident that V1 exhibits a more stable and energetically favorable binding profile. The overall binding energy for V1 (-240.37 kcal/mol) is significantly lower than that for V2 (-65.7268 kcal/mol), indicating a stronger interaction between V1 and TLR4. This difference is primarily driven by the higher contributions from lipophilic (-131.91 kcal/mol) and van der Waals interactions (-220.62 kcal/mol) in the V1-TLR4 complex, which are much greater than those observed for V2 (-17.0467 kcal/mol and −42.1463 kcal/mol, respectively).

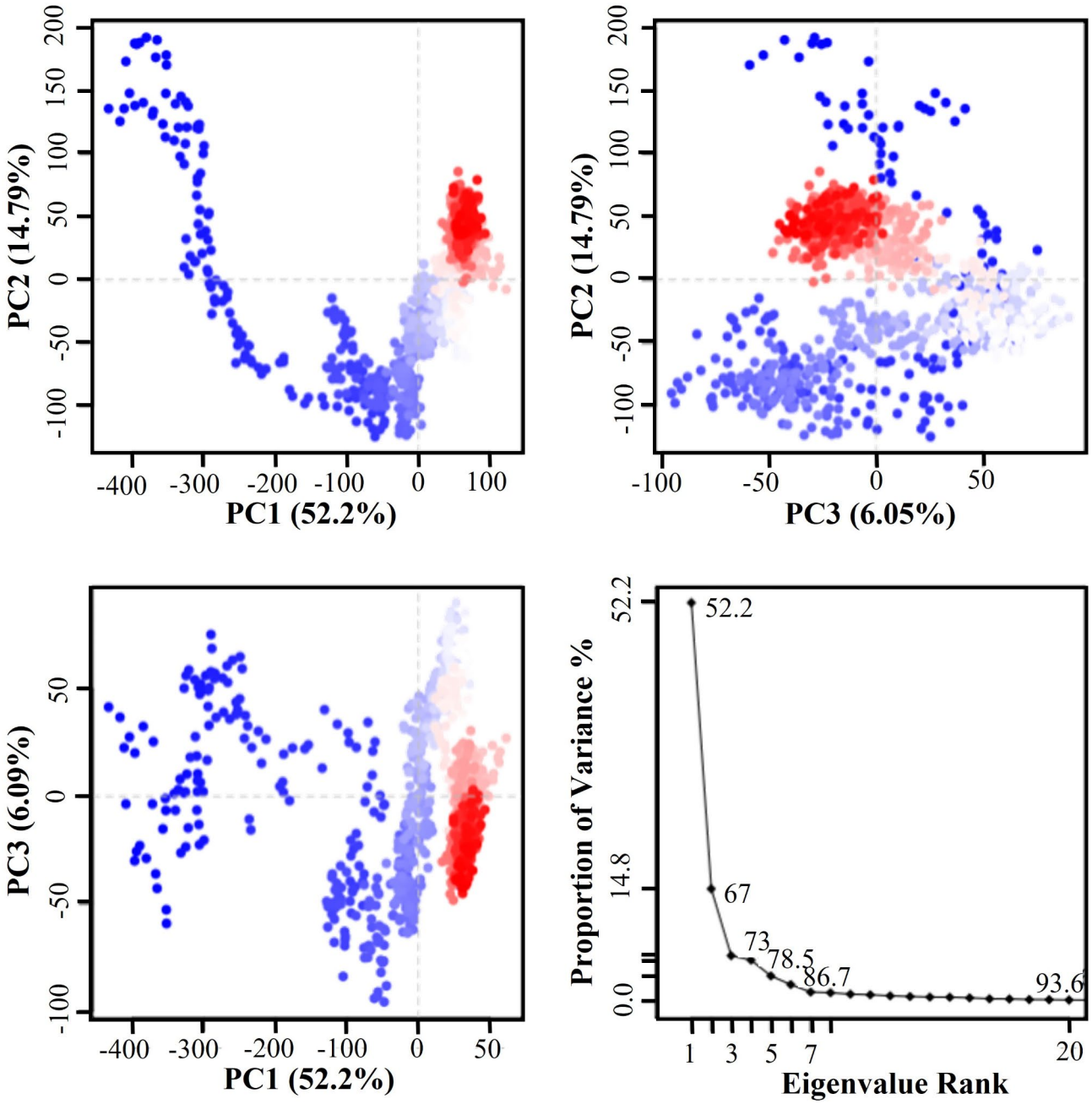

Fig. 11. PCA analysis of the V2-TLR4 complex performed by the R script.

Sr. No.	MM/GBSA variable	V1-TLR4	V2-TLR4
1	Bind	−240.366	−65.727
2	Bind_Lipo	−131.909	−17.047
3	Bind_vdW	−220.621	−42.146
4	Bind_Coulomb	115.280	−49.365
5	Bind_Hbond	−28.132	−5.0742
6	Bind_Packing	1.573	−3.5666

Table 5. Relative binding free energies (kcal/mol) between V1, V2 and TLR4.

The Coulombic interactions also differ notably between the two complexes. While V1 exhibits repulsive electrostatic interactions (115.28 kcal/mol), V2 shows attractive Coulombic interactions (-49.3653 kcal/mol). Despite this, the overall stability of the V1-TLR4 complex is still superior, suggesting that the strong lipophilic and van der Waals interactions in V1's favor outweigh the repulsive electrostatic forces. Hydrogen bonds further support the greater stability of V1, with a contribution of -28.13 kcal/mol compared to -5.07422 kcal/mol for V2. This indicates that more hydrogen bonds are formed in the V1-TLR4 complex, enhancing its stability. The packing interactions are relatively minor for both complexes, with similar contributions, indicating that both V1 and V2 fit well into the TLR4 binding pocket. Overall, V1 demonstrates a more stable and energetically favorable binding with TLR4 compared to V2, corroborating with the PCA analysis. The stronger lipophilic, van der Waals, and hydrogen bonding interactions in the V1-TLR4 complex are the primary contributors to this increased stability, despite the repulsive Coulombic interactions. These findings underscored the importance of various forces in determining the binding affinity and stability of ligand-receptor interactions.

Dynamic cross-correlation matrix

V1-TLR4 complex The dynamic cross-correlation matrix (DCCM) analysis of the V1-TLR4 docked complex over a 100 ns molecular dynamics (MD) simulation reveals significant insights into the structural dynamics and interactions within the complex. The DCCM exhibits distinct patterns of positive and negative correlations among the residues, indicated by a color scale ranging from −1.0 (green/yellow) to 1.0 (blue/dark blue) (Fig. 12A). Large blue blocks along the diagonal and off-diagonal regions suggest that certain domains or segments of the complex exhibit highly coordinated movements. Notably, regions around residues 500–700 and 1000–1200 display strong positive correlations, indicating that these areas move together during the simulation, potentially signifying functional or structurally interdependent domains. Conversely, smaller green and yellow regions denote areas where residues move in opposite directions, highlighting potential sites of mechanical coupling or functional antagonism. These anti-correlated motions, while less prevalent, are crucial for understanding the dynamic balance and functional interplay within the complex. The presence of both correlated and anti-correlated motions suggests a sophisticated level of structural orchestration necessary for the V1-TLR4 complex's functionality. Overall, the DCCM provided a comprehensive overview of the residue interactions and dynamic behavior, offering valuable insights for further functional and mechanistic studies of the V1-TLR4 docked complex.

V2-TLR4 complex The dynamic cross-correlation matrix (DCCM) analysis of the V2-TLR4-docked complex over a 100ns simulation reveals significant patterns of correlated motions among the residues (Fig. 12B). The DCCM demonstrates that regions of the protein exhibit both highly correlated (positive) and anti-correlated (negative) movements. Notably, the diagonal blocks of high positive correlation (dark blue) indicate that adjacent or nearby residues tend to move coherently, suggesting robust structural regions that maintain the integrity of the complex. Conversely, the scattered regions of negative correlation (yellow) indicate residues moving in opposite directions, which could signify compensatory adjustments within the protein structure. The predominance of neutral correlations (teal) suggests that a large portion of the residues move independently, reflecting the flexible nature of certain regions within the complex. Overall, these dynamic correlations suggest that the V2-TLR4 complex maintains a stable yet adaptable structure, capable of accommodating functional movements necessary for its biological activity.

Comparison of V1-TLR4 and V2-TLR4 complex The observed patterns in the DCCM indicate that the V2-TLR4-docked complex is relatively stable as compared to the V1-TLR4 complex. The presence of strong positive

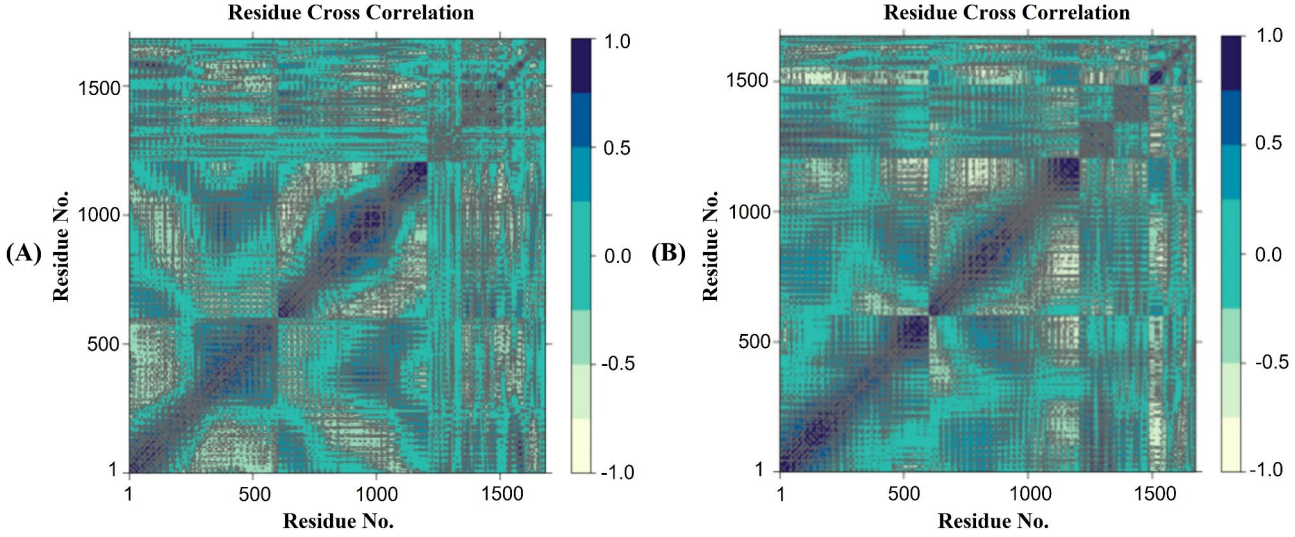

Fig. 12. DCCM analysis of the docked complex performed by the R script (**A**) V1-TLR4 (**B**) V2-TLR4.

correlations along the diagonal signifies that key structural regions move coherently, maintaining the complex's integrity. The existence of some anti-correlated motions does not detract from the overall stability but rather highlights dynamic flexibility that may be essential for functional regulation. The predominance of neutral correlations suggests that while certain regions are highly flexible, they do not compromise the overall stability of the complex. Therefore, the V2-TLR4 complex exhibits a balanced dynamic behavior, maintaining structural stability while allowing for necessary functional adaptability.

Conformational B-cell epitope prediction

The web server ElliPro was utilized to predict the conformational B-cell epitopes with values ranging from 0.508 to 0.961 for V1 and 0.545 to 0.984 for V2, respectively. Similarly, eight conformational B-cell epitopes were predicted for V1 and nine for V2 (Tables S5, S6 and Figs. S3, S4).

Immune simulation

The body's normal immunological responses to the modeled vaccine were determined by the C-ImmSim server. During immune simulations, the vaccines were administered in 3 doses at an interval of 1, 84, and 168 h (Fig. 13). After administering the vaccine injections, the number of activated antigen counts, B-cells, T_H-cells, and cytokines were increased (Figure S5 & S6). Antigen interaction initiates the first immune response, which mostly results in the production of IgM antibodies along with some IgG antibodies. Depending on the type of antigenic pathogen, minimal levels of antibodies are often produced. As the primary immunological response, the first dose of the vaccines (V1 & V2) caused an increase in IgM antibody levels. Two more antigen exposures spaced at different times resulted in the development of a strong immunological response marked by higher IgM and IgG levels. It was shown that the levels of IgM, IgG1, and IgG1 + IgG2 had significantly increased after the third dose. This indicated that the immunoglobulins were more likely to create immunological memory and had a higher binding affinity for the vaccine component (antigens). Different cells exhibited a robust response in the case of CD8 + and CD4 + T-cells, which was followed by the establishment of memory. Throughout the exposure, the frequency of T_H cells remained higher.

Codon adaptation and in-silico cloning

The Java Codon Adaptation Tool (JCat) was utilized to optimize vaccine constructs for expression in *Escherichia coli* (strain K12). By calculating Codon Adaptation Index (CAI) values, this tool ensured optimal codon usage by facilitating the back translation of vaccine amino acid sequences to DNA. The optimal codon adaptation was indicated by a CAI value of 0.985 and 0.971 for V1 and V2, respectively. The optimized sequence had a GC content of 48.83% for V1 and 48.32% for V2, suggesting a suitable expression probability in the *E. coli* expression system. The SnapGene software was used to add the vaccine-optimized sequence to the vector by using the NruI restriction site. After cloning, the vector and insert's ultimate length was found to be 7124 bp and 7097 bp for V1 and V2, respectively (Figure S7).

Identification of novel drug targets

The 31 cytoplasmic proteins were screened against the FDA-approved drug targets and it was found that none of the proteins showed similarity with the approved targets. Similarly, a characteristic analysis of the proteins was performed to find potential novel drug targets by figuring out their molecular weights, amino acid lengths, protein-protein interactions, and membrane-spanning helices. Because of solvent accessibility and ease of purification, proteins with molecular weights around 110 kDa, less than 800 amino acid length, and with a helix composition of 0 or 1 are thought to be the ideal prospects for therapeutic targets. The functional relationships among protein were aided by the STRING database. Proteins with an average node degree value (K) higher than five were identified as hub proteins (Fig. 14). After prioritizing and filtering proteins using the indicated tools, a total of eight novel drug targets including QBX69_06935, QBX69_03545, QBX69_00180, QBX69_01210, QBX69_00550, QBX69_00900, QBX69_00635, and QBX69_06170 were found and subjected to structural analysis.

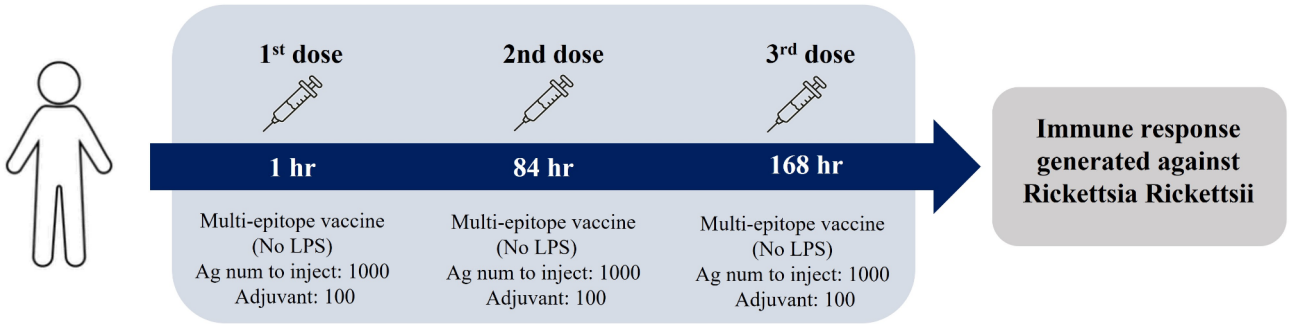

Fig. 13. Schematic representation of vaccination trial by C-ImmSim server at three doses 1 h, 84 h, 168 h produces an immune response against *R. rickettsii*.

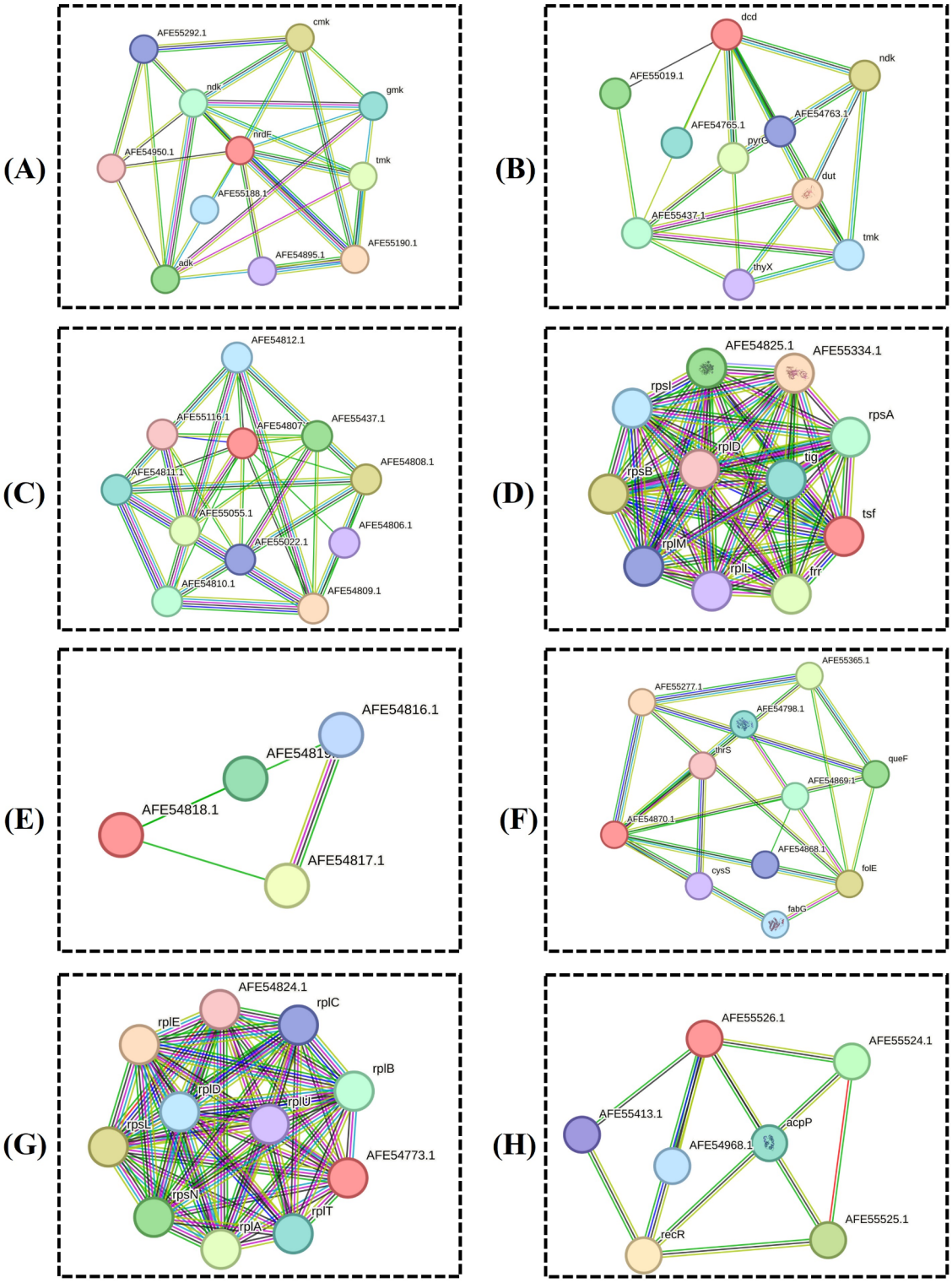

Fig. 14. Protein-protein interactions of the prioritized drug target proteins by the STRING database (**A**) QBX69_06935 (**B**) QBX69_03545 (**C**) QBX69_00180 (**D**) QBX69_01210 (**E**) QBX69_00550 (**F**) QBX69_00900 (**G**) QBX69_00635 (**H**) QBX69_06170.

Structure determination and drug pocket screening

The Swiss Model server was used to anticipate the targets' three-dimensional structures. Ramachandran plots validated these models by showing over 90% of the residues in favored regions (Table 6). Moreover, high ERRAT scores further confirmed the quality of the structures (Table 6). The drug-binding pockets for the prioritized

Sr. No.	Accession number	Protein description	ERRAT quality factor	Ramachandran favored (%)
1	QBX69_03545	DCTP deaminase	96.0396	92.00
2	QBX69_00550	Rhodanese-fold domain	92.6966	93.40
3	QBX69_00180	Protein-disulfide oxidoreductase	88.9401	87.50
4	QBX69_00635	ATP-binding cassette protein	98.3333	95.40
5	QBX69_00900	Lyase 6-Pyruvoyl tetrahydropterin synthase	89.1304	90.10
6	QBX69_01210	Elongation factor (EF)-tu protein	96.0317	92.90
7	QBX69_06170	DNA-binding protein EbfC/YbaB	97.7099	92.40
8	QBX69_06935	Ribonucleotide reductase (RNR)	96.9072	97.80

Table 6. Validation of the tertiary structures by ERRAT, Ramachandran score.

targets were determined by the DoGSiteScorer, and one protein (QBX69_00900) was excluded from the predicted novel targets due to a low druggable score (Fig. 15; Table 7).

Discussion

R. rickettsii is the cause of Rocky Mountain Spotted Fever (RMSF). It is the most common and serious rickettsial disorder in the United States. If medications are not timely administered then the death rate can reach 20–30% [2]. This pathogenic bacterium is equally harmful to humans and other species. It is a tick-borne disease spread by the bite of an infected tick. Ticks are considered to be a vector and reservoir of the *R. rickettsii*[3]. Further, there is no potential treatment available against this harmful bacterium due to non-specific symptoms and delayed treatment. Therefore, the objective of this study was to use the core proteome of *R. rickettsia* to identify potential therapeutic targets and construct a broad-spectrum multi-epitope vaccine that can stimulate the host immune system and produce innate and adaptive immunity. The core proteome was employed because the core proteins are present in all of the strains of the targeted pathogen, these proteins are becoming increasingly important and, when included in therapeutics, provide immune protection against a broader range of species[64]. As a result, following a rigorous analysis, four potential vaccines and seven novel drug target candidate proteins were obtained.

The vaccine targets proteins are ideal targets because they were found to be in the extracellular and membrane domains, and they might be the initial molecules that interact with the host cells[64]. They are easily accessible and interact with host cells, eliciting potent B-cell and T-cell responses[65]. Strict criteria were used to identify B- and T-cell binding epitopes[38], as they produce cell-mediated immunization, helper T-cells induce B-cells to release antibodies, macrophages to phagocytose pathogens, and cytotoxic T-cells to destroy specific parasitized cells[66] and B-cell epitopes mediated humoral immunity by producing antibodies against antigens[37]. To prioritize T- and B-cell epitopes, several factors such as non-toxicity, non-allergenicity, antigenicity, virulence, solubility, immunogenicity, and cytokine induction have been analyzed and considered. Top-ranked T- and B-cell binding epitopes, their linkers, and the adjuvant sequences were utilized in the construction of the two vaccine constructs; V1 & V2. An antigenic protein called HBHA has been linked to bacterial adhesion to epithelial cells and widespread dissemination[67]. It is a surface-displayed protein that serves as an adhesin for non-phagocytic cells that can affect binding to host cells via a unique, methylated, carboxyl-terminal, lysine-, alanine-, and proline-rich repeat region[68]. As an effective immune adjuvant, HBHA can induce a strong T_{H1} cell immune response and plays an important role in the research of multi-epitope vaccines for immunotherapy (i.e., tumor vaccines). Recently, heparin-binding hemagglutinin (HBHA), a component of *Mycobacterium tuberculosis*, has been closely investigated for its strong immune potential, which can stimulate the migration of DCs and promote the expression of a variety of surface molecules (e.g., CD40, CD80, and CD86), MHC I and MHC II molecules, as well as inflammatory cytokines (e.g., IL-6, IL-12, IL-1β, and TNF-α) in a TLR4-dependent manner[69]. It was anticipated that the proposed vaccine constructs might minimize toxicity and allergenicity while exhibiting strong antigenicity values. Physicochemical analysis predicted that the designed constructs are capable of inducing a strong immune response against *R. rickettsii*. These vaccines exhibited small molecular weight, good solubility, and thermodynamic stability, which qualify them for large-scale production and delivery. The secondary and tertiary structures were formulated and validated. The Ramachandran plot is an important parameter for the validation of the 3D structures of the proteins. Based on steric hindrance, this plot assessment describes in detail the arrangement and direction of dihedral angles that fall within the forbidden area. Residues in the Ramachandran plot's most preferred areas show improved protein structure folding and support the structure's quality[70].

For the effectiveness of a vaccine, it is important to evaluate its interaction pattern with the host immune cells. TLR4 is an essential component of the human immune system and a member of the pattern recognition receptor family, which is extremely selective and sensitive in its response to invasive infections. It is responsive to pathogen-associated lipopolysaccharides (LPS) and lipo-oligosaccharides and capable of recognizing pathogen-associated molecular patterns (PAMPs) from different pathogens. In addition to PAMPs, some endogenous ligands produced as a result of inflammation or tissue injury can also activate TLR4. This receptor-ligand interaction initiates an intracellular signaling cascade that in turn causes the subsequent pro-inflammatory response. TLR4 is considered a potential therapeutic target because it is implicated in several clinical diseases[63]. Therefore, here we used this receptor to evaluate the interacting potential of our designed vaccine constructs. The docking analysis indicated that there are significant binding affinity and interactions between the vaccines

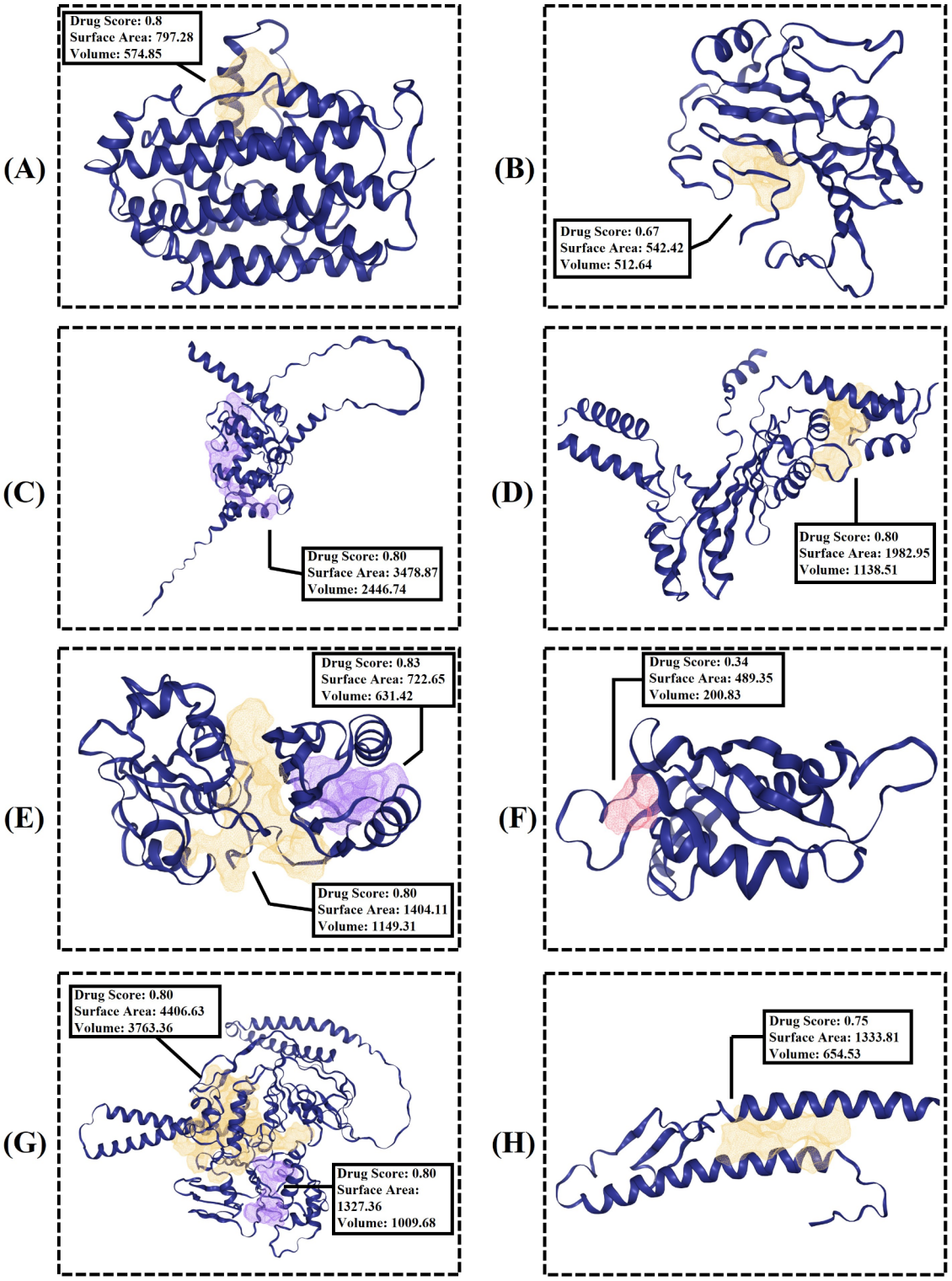

Fig. 15. Drug-pocket screening of the novel drug targets by the DoGsiteScorer (**A**) QBX69_06935 (**B**) QBX69_03545 (**C**) QBX69_00180 (**D**) QBX69_01210 (**E**) QBX69_00550 (**F**) QBX69_00900 (**G**) QBX69_00635 (**H**) QBX69_06170.

and TLR4 [71]. The results of the docking were verified by various analyses. The analysis included the MD data (RMSD and RMSF). Moreover, PCA, binding energy calculation through MM-GBSA and DCCM analysis were also conducted, which confirmed the in-silico immunogenic interaction of epitopes. Among these analyses, the MM-GBSA evaluation identified the most favorable and lowest energy conformations in the TLR4-vaccine complexes. Understanding the complex's dynamic behavior and structure was made possible by the PCA analysis

Accession No.	Pocket	Drugability	Residues	Volume	Surface
QBX69_03545	P_0	0.791452	THR-24,ILE-28,TRP-30,ARG-190,VAL-189,TRP-187,ALA-186,LEU-289,VAL-292,GLU-293,ASN-296,PHE-297,PHE-318,THR-319,GLU-320	574.85	797.28
QBX69_00550	P_0	0.668234	TYR-177,PRO-186,LEU-185,GLN-180,THR-184,VAL-181,VAL-183,GLY-182,GLU-137,LEU-136,ARG-93,THR-135,VAL-134,HIS-133,GLU-127,LEU-126,THR-124,VAL-123	512.64	542.42
QBX69_00180	P_0	0.802788	ALA-186,THR-196,GLY-199,GLN-200,SER-205,PRO-206,GLU-207,TYR-209LYS-210,LEU-213,ASN-214,ASN-215,ASP-216,THR-219,ALA-220,ASN-227,ALA-230,LYS-231,ALA-223,PRO-233,GLN-234,PHE-235,GLY-237,THR-238,PRO-239,SER-240,PHE-241,THR-248,GLY-249,SER-250,TYR-251,ILE-150,ALA-151,THR-152,LYS-153,GLN-48,ASN-47,LYS-44,TYR-43,GLU-41,ASN-40,ASN-39,GLN-38,THR-37,HIS-35,GLU-34,ILE-33,GLN-31,LYS-30,HIS-27,LYS-26,GLN-23,CYS-118	2446.74	3478.87
	P_1	0.808775	ARG-163,CYS-164,LYS-165,GLY-166,ASP-167,SER-170,PHE-171,GLN-173,LEU-174,GLN-181,TRP-185,LEU-194,LEU-195,ASP-197,ILE-198,LEU-201,GLY-202,GLY-203,LYS-217,ILE-218,ILE-221,LEU-222,ALA-224,ASN-225,THR-91,PHE-92,LYS-93,VAL-94,ASN-97,ASP-98,MET-99,VAL-100,ILE-109,VAL-110,VAL-111,GLY-245,ASN-244,VAL-243,PHE-242,	1138.51	1982.95
QBX69_00635	P_0	0.830064	ALA-28,LYS-27,ASP-24,GLY-21,ALA-20,GLY-19,THR-18,LYS-17,ILE-40,LEU-43,ARG-44,LYS-46,GLY-47,LEU-48,THR-132,ALA-131,ASN-129,GLU-128,ILE-127,GLU-125,GLU-124,GLN-117,GLN-90,ASP-93,LEU-94	864.88	1351.08
QBX69_00900	P_0	0.804956	GLN-208,LEU-207,GLY-204,ILE-7.ILE-8,SER-9,ALA-10,TYR-11,THR-179,GLY-180,GLY-181,ILE-182,ASN-135,ARG-134,THR-133,THR-152,PHE-155,THR-154,GLN-157,PHE-158,ASN-229,ASP-228,PHE-227,VAL-226,LYS-76,ILE-177,ASN-78,PHE-85,GLN-86,LYS-87,LYS-89,ALA-98	1149.31	1404.11
	P_1	0.838894	GLY-69,LEU-67,LEU-64,GLU-63,VAAL-60,LYS-89,VAL-90,LEU-8,SER-9,ALA-10,TYR-11,ILE-16,GLU-18,LEU-22,ILE-23,LYS-25,LEU-26,LEU-27,LEU-29,GLY-30,GLY-37,THR-38,ILE-39,PHE-46,ASN-47,GLY-48,SER-49	631.42	722.65
QBX69_01210	P_3	0.343427	HIS-106,THR-105,ASN-102,ASN-96,GLN-95,MET-94	200.83	489.35
QBX69_06170	P_0	0.80635	SER-327,LYS-326,GLY-325,SER-324,GLY-323,ASN-322,ASP-152,LEU-151,HIS-150,ASN-149,THR-148,SER-464,ASP-466,TYR-456,LEU-453,ILE-452,ASP-451,LEU-449,ILE-448,GLU-447,ASP,444,MET-443,ASN-439,ASP-442,ALA-219,LEU-220,GLU-217,GLU-216,ILE-213,ILE-212,ASP-206,PHE-205,PHE-204,LYS-203,SER-181,THR-178,ARG-177,ASP-176,HIS-175,SER-174,VAL-173,ASN-400,PHE-399,MET-398,GLN-396,LYS-395,GLY-392,MET-426,LEU-425	3763.36	4406.43
QBX69_06935	P_1	0.806491	LYS-79,ILE-78,PRO-77,ILE-76,GLN-74,ARG-73,LEU-72,TYR-71GLY-70,THR-69,SER-51,LEU-159,ASN-149,THR-148,PRO-147,GLU-146,PRO-139,LEU-137,ILE-136LEU-135,ALA-134,LEU-131	1009.68	1327.36
	P_0	0.751011	GLN-10,SER-13,MET-14,LYS-17,MET-18,ALA-21,GLN-22,GLN-24,MET-25,ALA-26,MET-97,GLY-96,LEU-94,ALA-93,LEU-90,SER-89,SER-86,ASP-83,CYS-82,GLY-45	654.53	1333.81

Table 7. Druggability analysis of the prioritized drug target proteins.

of the V1 and V2 docked complexes. Based on the several clusters seen in the PCA plots, the findings showed that the V1-TLR4 and V2-TLR4 complex display multiple stable conformations. For additional functional and mechanistic investigations of the V1-TLR4 and V2-TLR4 docked complexes, the DCCM offered a thorough summary of the residue interactions and dynamic behavior. Immune simulation analysis showed reliable and effective immune responses. Repetitive exposure to the antigenic complex V1 and V2 strengthened immune responses. To shield the host from *R. rickettsii* exposure in the future, memory T- and B-cells as well as helper and cytotoxic T cells were generated. The immunological humoral response was enhanced by TH cell and IgG production which was increasing with each dose[72]. The potential of vaccines for the expression in the bacterial expression system was ensured by the in-silico cloning of the optimized vaccine in an *E. coli* plasmid[73]. However, this study has some limitations. The degree of protection against *R. rickettsii* is unknown, and the accuracy of these methods may be limited. Even though immunoinformatics predictions have been implicated in recent promising case reports[28,74,75], more in-vitro and in-vivo studies are required to validate the findings of this study.

Of the seven novel drug target proteins, ribonucleotide reductase (RNR) (QBX69_06935) converts ribonucleotides to deoxyribonucleotides, necessary for DNA replication and repair and thus making it an important target for anti-proliferative drugs against bacteria, viruses, cancer, and parasites[76]. The DCTP deaminase (QBX69_03545) is a significant drug target because it catalyzes the conversion of dCTP to dUTP, necessary for DNA replication. By emphasizing this enzyme, treatments might be possible as it may interfere with the production of nucleotides required for the survival and replication of this obligate intracytoplasmic bacteria[77]. A protein-disulfide oxidoreductase (QBX69_00180) that is encoded by DsbA catalyzes the creation of disulfide bonds in the periplasm during the folding of released proteins. It is hypothesized that both rickettsial DsbA proteins, one through a transmembrane domain and the other as a lipoprotein, are membrane-anchored and crucial for a wide variety of virulence factors[78]. Another drug target protein is the elongation factor (EF)-tu protein (QBX69_01210). This cytoplasmic protein that delivers and catalyzes the aminoacyl-tRNA binding to the ribosome is crucial for protein synthesis. But in addition to its primary role, EF-tu can bind fibronectin and plasminogen, respectively, to act as an adhesin on the surface of several pathogens, including *Mycobacterium pneumonia* and *Pseudomonas aeruginosa*[79]. The rhodanese-fold domain (QBX69_00550) is essential for a variety of physiological processes, including cell cycle regulation, sulfur transport, and hydrogen sulfide metabolism. Rhodanese domains that are catalytically active and possess a cysteine residue are versatile as they can also be involved in tRNA modification and hydroxylation[80]. The lyase 6-Pyruvoyl tetrahydropterin synthase (6-PTPS) (QBX69_00900) is involved in the production of tetrahydrobiopterin. It functions in the folate biosynthesis pathway required for the growth and survival of the pathogen in which dihydroneopterin aldolase (DHNA) is lacking. It has been identified as a potential drug target[81]. ATP-binding cassette proteins (QBX69_00635) that target the ribosome of Gram-positive pathogens contain members of the ABC-F subfamily, which mediates resistance to a wide range of therapeutically relevant antibiotic classes. The effectiveness of existing antibiotics can be restored by inhibiting these proteins and have drawn attention as possible targets for drugs[82]. The DNA-binding protein EbfC/YbaB (QBX69_06170) colocalizes with nucleoids and controls the production of surface

lipoproteins associated with virulence. It binds DNA both specifically and non-specifically. While binding DNA, it creates homodimers and other ordered structures[83]. The druggability analysis indicates that the seven novel drug target proteins have the potential to be experimentally validated for the discovery of alternative therapeutic targets to combat this deadly bacterium.

Conclusion

Before conducting biological studies, it is critical to rely on in-silico methods to provide a more accurate understanding of the likelihood and viability of identifying new targets for drugs and vaccines. One possible strategy to combat *R. rickettsii* is the use of a proteomics-based technique to find potential therapeutic targets. The foundation for the potential therapeutic target identification was laid by the thorough examination of the bacterial proteomes of 17 strains, which identified seven novel drug and four novel vaccine targets. To identify the most promising drug candidates, several biological databases, comparative sequencing analysis, and druggability assessments were carried out. In the end, seven proteins were shown as novel druggable candidates to be studied as potential targets against the bacterium. Furthermore, using immunoinformatics, molecular docking, MD simulation techniques, and various important analyses such as MM-GBSA, PCA, and DCCM analysis, two safe and latent vaccine constructs that may elicit humoral, cellular, and innate immune responses were designed. Such vaccines might provide wide protection against different strains of *R. rickettsii* by addressing antigenic diversity and enhancing immune recognition by targeting several epitopes. In addition to enhancing public health outcomes, this innovative approach holds considerable promise for addressing the growing danger of antibiotic resistance. To develop the potential therapeutics for this bacterium, the drug targets and the vaccine construct designed in this study need to be validated through experiments.

Data availability

Data availability All the data analyzed in this work is present in this Manuscript and the in the supplementary materials.

Received: 21 August 2024; Accepted: 13 December 2024
Published online: 06 January 2025

References

1. Helminiak, L., Mishra, S. & Kim, H. K. Pathogenicity and virulence of <i> Rickettsia. *Virulence* **13**, 1752–1771 (2022).
2. Parola, P. et al. Update on Tick-Borne Rickettsioses around the World: a Geographic Approach. *Clin. Microbiol. Rev.* **26**, 657–702 (2013).
3. Azad, A. Rickettsial Pathogens and Their Arthropod Vectors. *Emerg. Infect. Dis.* **4**, 179–186 (1998).
4. Zhang, Y. Y. et al. Mapping the global distribution of spotted fever group rickettsiae: a systematic review with modelling analysis. *Lancet Digit. Health* **5**, e5–e15 (2023).
5. Yaglom, H. D., Nicholson, W. L., Casal, M., Nieto, N. C. & Adams, L. Serologic assessment for exposure to spotted fever group rickettsiae in dogs in the Arizona–Sonora border region. *Zoonoses Public. Health* **65**, 984–992 (2018).
6. Álvarez-López, D. I. et al. Epidemiology and Clinical Features of Rocky Mountain Spotted Fever from Enhanced Surveillance, Sonora, Mexico: 2015–2018. *Am. J. Trop. Med. Hyg.* **104**, 190–197 (2021).
7. Hidalgo, M. et al. Outbreak of Rocky Mountain spotted fever in Córdoba, Colombia. *Mem. Inst. Oswaldo Cruz* **106**, 117–118 (2011).
8. Estrada-Mendizabal, R. J. et al. Rickettsial Disease Outbreak, Mexico, *Emerg Infect Dis* 29, (2023). (2022).
9. Chung, I. H., Robinson, L. K., Stewart-Juba, J. J., Dasch, G. A. & Kato, C. Y. Analytically Sensitive Rickettsia Species Detection for Laboratory Diagnosis. *Am. J. Trop. Med. Hyg.* **106**, 1352–1357 (2022).
10. Biggs, H. M. et al. Diagnosis and Management of Tickborne Rickettsial Diseases: Rocky Mountain Spotted Fever and Other Spotted Fever Group Rickettsioses, Ehrlichioses, and Anaplasmosis — United States. *MMWR Recommendations Rep.* **65**, 1–44 (2016).
11. Binder, A. M. & Armstrong, P. A. Patient characteristics, treatment patterns, and outcomes of Rickettsial diseases among a commercially insured population in the United States, 2005–2017. *Sci. Rep.* **11**, 18382 (2021).
12. Wormser, G. P., Strle, F. & Shapiro, E. D. Is Doxycycline Appropriate for Routine Treatment of Young Children With Erythema Migrans? *Pediatr. Infect. Disease J.* **38**, 1113–1114 (2019).
13. Alhassan, A. et al. Rickettsia rickettsii Whole-Cell Antigens Offer Protection against Rocky Mountain Spotted Fever in the Canine Host. *Infect. Immun.* **87**, (2019).
14. Blom, J. et al. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res.* **44**, W22–W28 (2016).
15. Sethi, G., Sethi, S. & Krishna, R. Multi-epitope based vaccine design against Staphylococcus epidermidis: A subtractive proteomics and immunoinformatics approach. *Microb. Pathog* **165**, 105484 (2022).
16. Apweiler, R. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* 32, 115D – 119 (2004).
17. Dey, J. et al. Molecular Characterization and Designing of a Novel Multiepitope Vaccine Construct Against Pseudomonas aeruginosa. *Int. J. Pept. Res. Ther.* **28**, 1–19 (2022).
18. Kumar, A., Thotakura, P. L., Tiwary, B. K. & Krishna, R. Target identification in Fusobacterium nucleatum by subtractive genomics approach and enrichment analysis of host-pathogen protein-protein interactions. *BMC Microbiol.* **16**, 84 (2016).
19. Mursaleen, S. et al. Genome-level therapeutic targets identification and chimeric Vaccine designing against the Blastomyces dermatitidis. *Heliyon* **10**, e36153 (2024).
20. Sah, P. P., Bhattacharya, S., Banerjee, A. & Ray, S. Identification of novel therapeutic target and epitopes through proteome mining from essential hypothetical proteins in Salmonella strains: An In silico approach towards antivirulence therapy and vaccine development. *Infect. Genet. Evol.* **83**, 104315 (2020).
21. Shah, M. et al. Delineating Novel Therapeutic Drug and Vaccine Targets for Staphylococcus cornubiensis NW1T Through Computational Analysis. *Int. J. Pept. Res. Ther.* **27**, 181–195 (2021).
22. Irudal, S. et al. Identification by Reverse Vaccinology of Three Virulence Factors in Burkholderia cenocepacia That May Represent Ideal Vaccine Antigens. *Vaccines (Basel)* **11**, 1039 (2023).
23. Hizbullah et al. Reverse vaccinology and subtractive genomics-based putative vaccine targets identification for Burkholderia pseudomallei Bp1651. *Microb. Pathog* **125**, 219–229 (2018).

24. Gupta, S. K. et al. ARG-ANNOT, a New Bioinformatic Tool To Discover Antibiotic Resistance Genes in Bacterial Genomes. *Antimicrob. Agents Chemother.* **58**, 212–220 (2014).
25. Ammari, M. G., Gresham, C. R., McCarthy, F. M. & Nanduri, B. HPIDB 2.0: a curated database for host–pathogen interactions. *Database* baw103 (2016). (2016).
26. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185 (2007).
27. Prabhu, D. et al. In silico Functional Annotation and Characterization of Hypothetical Proteins from Serratia marcescens FGI94. *Biology Bull.* **47**, 319–331 (2020).
28. Aslam, M. et al. Chlamydia trachomatis core genome data mining for promising novel drug targets and chimeric vaccine candidates identification. *Comput. Biol. Med.* **136**, 104701 (2021).
29. Doytchinova, I. A. & Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **8**, 4 (2007).
30. Dimitrov, I., Bangov, I., Flower, D. R. & Doytchinova, I. AllerTOP v.2—a server for in silico prediction of allergens. *J. Mol. Model.* **20**, 2278 (2014).
31. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *J. Mol. Biol.* **305**, 567–580 (2001).
32. Cash, P. Protein identification and analysis tools in the ExPASy server. *Cell. Biol. Int.* **23**, 385 (1999).
33. Sarfraz, A. et al. Structural informatics approach for designing an epitope-based vaccine against the brain-eating Naegleria fowleri. *Front. Immunol.* **14**, 1–17 (2023).
34. Vita, R. et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* **43**, D405–D412 (2015).
35. Qasim, A. et al. Computer-aided genomic data analysis of drug-resistant Neisseria gonorrhoeae for the Identification of alternative therapeutic targets. *Front. Cell. Infect. Microbiol.* **13**, 1017315 (2023).
36. Halary, F. et al. Control of self-reactive cytotoxic T lymphocytes expressing γδ T cell receptors by natural killer inhibitory receptors. *Eur. J. Immunol.* **27**, 2812–2821 (1997).
37. Hohman, L. S. & Peters, N. C. CD4 + T Cell-Mediated Immunity against the Phagosomal Pathogen Leishmania: Implications for Vaccination. *Trends Parasitol.* **35**, 423–435 (2019).
38. Jaan, S. et al. Multi-epitope chimeric vaccine designing and novel drug targets prioritization against multi-drug resistant Staphylococcus pseudintermedius. *Front. Microbiol.* **13**, (2022).
39. Aiman, S. et al. Immunoinformatic-guided novel mRNA vaccine designing to elicit immunogenic responses against the endemic Monkeypox virus. *J. Biomol. Struct. Dyn.* **1–15**. https://doi.org/10.1080/07391102.2023.2233627 (2023).
40. Kakakhel, S. et al. Annotation of Potential Vaccine Targets and Designing of mRNA-Based Multi-Epitope Vaccine against Lumpy Skin Disease Virus via Reverse Vaccinology and Agent-Based Modeling. *Bioengineering* **10**, 430 (2023).
41. Magnan, C. N., Randall, A., Baldi, P. & SOLpro Accurate sequence-based prediction of protein solubility. *Bioinformatics* **25**, 2200–2207 (2009).
42. Jaan, S., Zaman, A., Ahmed, S., Shah, M. & Ojha, S. C. mRNA Vaccine Designing Using Chikungunya Virus E Glycoprotein through Immunoinformatics-Guided Approaches. *Vaccines (Basel)* **10**, (2022).
43. Geourjon, C. & Deléage, G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* **11**, 681–684 (1995).
44. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
45. Waterhouse, A. et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
46. Heo, L., Park, H. & Seok, C. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* **41**, W384–W388 (2013).
47. Shah, M. et al. Deciphering the Immunogenicity of Monkeypox Proteins for Designing the Potential mRNA Vaccine. *ACS Omega* **8**, 43341–43355 (2023).
48. Kozakov, D. et al. The ClusPro web server for protein-protein docking. *Nat. Protoc.* **12**, 255–278 (2017).
49. Prasanth, D. S. N. B. K. et al. In silico identification of potential inhibitors from Cinnamon against main protease and spike glycoprotein of SARS CoV-2. *J. Biomol. Struct. Dyn.* **39**, 4618–4632 (2021).
50. Shivakumar, D. et al. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **6**, 1509–1519 (2010).
51. Kerrigan, J. E. Molecular Dynamics Simulations in Drug Design. in 95–113 doi: https://doi.org/10.1007/978-1-62703-342-8_7 (2013).
52. Bera, I. & Payghan, P. V. Use of Molecular Dynamics Simulations in Structure-Based Drug Discovery. *Curr. Pharm. Des.* **25**, 3339–3349 (2019).
53. Bowers, K. J. et al. Molecular dynamics—Scalable algorithms for molecular dynamics simulations on commodity clusters. in *Proceedings of the ACM/IEEE conference on Supercomputing - SC '06* 84 (ACM Press, New York, New York, USA, 2006). doi: https://doi.org/10.1145/1188455.1188544 (2006).
54. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696 (2006).
55. Alandijany, T. A. et al. A multi-targeted computational drug discovery approach for repurposing tetracyclines against monkeypox virus. *Sci. Rep.* **13**, 14570 (2023).
56. Ponomarenko, J. et al. ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinform.* **9**, 1–8 (2008).
57. Rapin, N., Lund, O. & Castiglione, F. Immune system simulation online. *Bioinformatics* **27**, 2013–2014 (2011).
58. Grote, A. et al. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* **33**, 526–531 (2005).
59. Rahman, N., Shah, M., Muhammad, I., Khan, H. & Imran, M. Genome-wide Core Proteome Analysis of Brucella melitensis Strains for Potential Drug Target Prediction. *Mini-Reviews Med. Chem.* **21**, 2778–2787 (2020).
60. Qureshi, N. A. et al. Genome-Based Drug Target Identification in Human Pathogen Streptococcus gallolyticus. *Front. Genet.* **12**, 1–20 (2021).
61. Szklarczyk, D. et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
62. Shah, M. et al. Proteome level analysis of drug-resistant Prevotella melaninogenica for the identification of novel therapeutic candidates. *Front. Microbiol.* **14**, (2023).
63. Sarfraz, A. et al. Decrypting the multi-genome data for chimeric vaccine designing against the antibiotic resistant Yersinia pestis. *Int. Immunopharmacol.* **132**, 111952 (2024).
64. Gouda, A. M. et al. Integration of immunoinformatics and cheminformatics to design and evaluate a multitope vaccine against Klebsiella pneumoniae and Pseudomonas aeruginosa coinfection. *Front. Mol. Biosci.* **10**, (2023).
65. Wang, Y., Yang, K. & Zhou, H. Immunogenic proteins and potential delivery platforms for mpox virus vaccine development: A rapid review. *Int. J. Biol. Macromol.* **245**, 125515 (2023).
66. Todryk, S. M. T Cell Memory to Vaccination. *Vaccines (Basel)* **6**, 84 (2018).
67. de Lima, C. S. et al. Heparin-binding hemagglutinin (HBHA) of Mycobacterium leprae is expressed during infection and enhances bacterial adherence to epithelial cells. *FEMS Microbiol. Lett.* **292**, 162–169 (2009).

68. Parra, M. et al. The Mycobacterial Heparin-Binding Hemagglutinin Is a Protective Antigen in the Mouse Aerosol Challenge Model of Tuberculosis. *Infect. Immun.* **72**, 6799–6805 (2004).
69. Lei, Y. et al. Application of built-in adjuvants for epitope-based vaccines. 1–48 doi: https://doi.org/10.7717/peerj.6185 (2019).
70. Ho, B. K. & Brasseur, R. The Ramachandran plots of glycine and pre-proline. *BMC Struct. Biol.* **5**, 14 (2005).
71. Abraham, M. J. et al. High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX 1–2* **GROMACS**, 19–25 (2015).
72. Song, X., Zhang, H., Zhang, D., Xie, W. & Zhao, G. Bioinformatics analysis and epitope screening of a potential vaccine antigen TolB from Acinetobacter baumannii outer membrane protein. *Infect. Genet. Evol.* **62**, 73–79 (2018).
73. Dey, J. et al. Designing of multi-epitope peptide vaccine against Acinetobacter baumannii through combined immunoinformatics and protein interaction–based approaches. *Immunol. Res.* **71**, 639–662 (2023).
74. Farzan, M. et al. Immunoinformatics-based multi-epitope vaccine design for the re-emerging monkeypox virus. *Int. Immunopharmacol.* **123**, 110725 (2023).
75. Aslam, M. et al. Potential druggable proteins and chimeric vaccine construct prioritization against Brucella melitensis from species core genome data. *Genomics* **112**, 1734–1745 (2020).
76. Torrents, E. Ribonucleotide reductases: essential enzymes for bacterial life. *Front. Cell. Infect. Microbiol.* **4** (2014).
77. Speed, R. R. & Winkler, H. H. Deamination of deoxycytidine nucleotides by the obligate intracytoplasmic bacterium Rickettsia prowazekii. *J. Bacteriol.* **173**, 4902–4903 (1991).
78. Felsheim, R. F., Kurtti, T. J. & Munderloh, U. G. Genome Sequence of the Endosymbiont Rickettsia peacockii and Comparison with Virulent Rickettsia rickettsii: Identification of Virulence Factors. *PLoS One* **4**, e8361 (2009).
79. Matos, A. L., Curto, P. & Simões, I. Moonlighting in Rickettsiales: Expanding Virulence Landscape. *Trop. Med. Infect. Dis.* **7**, 32 (2022).
80. Alsohaibani, R. et al. Rhodanese-Fold Containing Proteins in Humans: Not Just Key Players in Sulfur Trafficking. *Antioxidants* **12**, 843 (2023).
81. Chinedu, S. N., Bella-Omunagbe, M., Okafor, E., Afolabi, R. & Adebiyi, E. Computational Studies on 6-Pyruvoyl Tetrahydropterin Synthase (6-PTPS) of *Plasmodium falciparum*. *Bioinform Biol. Insights* **18** (2024).
82. Sharkey, L. K. R., Edwards, T. A. & O'Neill, A. J. ABC-F Proteins Mediate Antibiotic Resistance through Ribosomal Protection. *mBio* 7, (2016).
83. Cordeiro, T. F. V. B., Gontijo, M. T. P., Jorge, G. P. & Brocchi, M. EbfC/YbaB: A Widely Distributed Nucleoid-Associated Protein in Prokaryotes. *Microorganisms* **10**, 1945 (2022).

Acknowledgements

Authors wish to thanks Researchers Supporting Project Number (RSP2025R110) at King Saud University Riyadh Saudi Arabia for financial support. The authors are also thankful to the Bahauddin Zakariya University, Multan, Pakistan for providing the necessary infrastructure to perform this research.

Author contributions

Fizza Arshad: Investigation, Visualization, Formal analysis, Writing- Original draft preparation; Asifa Sarfraz: Methodology, Formal analysis, Writing- Original draft preparation. Muhammad Shehroz: Investigation, Visualization, Writing- Original draft preparation; Umar Nishan: Investigation, Writing- Reviewing and Editing; Asia Perveen: Formal analysis, Validation, Writing- Reviewing and Editing; Riaz Ullah: Investigation, Visualization; Mohamed A. Ibrahim: Funding acquisition, Writing- Reviewing and Editing; Mohibullah Shah: Conceptualization, Supervision, Funding acquisition, Writing- Reviewing and Editing.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-83395-3.

Correspondence and requests for materials should be addressed to M.S.