Original Research Article

Development of multi-epitope mRNA vaccine against *Clostridioides difficile* using reverse vaccinology and immunoinformatics approaches

Caixia Tan [a,b], Yuanyuan xiao [a,b], Ting Liu [a,b], Siyao Chen [a,b], Juan Zhou [a,b], Sisi Zhang [a,b], Yiran Hu [a,b], Anhua Wu [a,b,**], Chunhui Li [a,b,*]

[a] *Infection Control Center, Xiangya Hospital, Central South University, Changsha, Hunan Province, 410008, China*
[b] *National Clinical Research Center for Geriatric Disorders (XiangYa Hospital), Changsha, Hunan Province, 410008, China*

ABSTRACT

Clostridioides difficile (*C. difficile*), as the major pathogen of diarrhea in healthcare settings, has become increasingly prevalent within community populations, resulting in significant morbidity and mortality. However, the therapeutic options for *Clostridioides difficile* infection (CDI) remain limited, and as of now, no authorized vaccine is available to combat this disease. Therefore, the development of a novel vaccine against *C. difficile* is of paramount importance. In our study, the complete proteome sequences of 118 strains of *C. difficile* were downloaded and analyzed. We found four antigenic proteins that were highly conserved and can be used for epitope identification. We designed two vaccines, WLcd1 and WLcd2, that contain the ideal T-cell and B-cell epitopes, adjuvants, and the pan HLA DR-binding epitope (PADRE) sequences. The biophysical and chemical assessments of these vaccine candidates indicated that they were suitable for immunogenic applications. Molecular docking analyses revealed that WLcd1 bonded with higher affinity to Toll-like receptors (TLRs) than WLcd2. Furthermore, molecular dynamics (MD) simulations, performed using Gmx_MMPBSA v1.56, confirmed the binding stability of WLcd1 with TLR2 and TLR4. The preliminary findings suggested that this multi-epitope vaccine could be a promising candidate for protection against CDI; however, experimental studies are necessary to confirm these predictions.

1. Introduction

Clostridioides difficile (*C. difficile*), a gram-positive anaerobic bacterium, is the main cause of antibiotic-associated diarrhea (AAD) and a significant contributor to fatalities associated with gastrointestinal illness in both North American and European healthcare settings [1]. It poses a significant healthcare challenge worldwide due to its high morbidity and mortality rates, alongside its tendency to recur and the emerging issue of antibiotic resistance [2–4]. While antibiotics exist to treat primary *Clostridioides difficile* infections (CDI), their effectiveness is compromised by challenges such as recurrent infections and the emergence of antibiotic-resistant strains, continuing to pose a significant threat [5,6]. Consequently, it is imperative to explore new and alternative strategies for the prevention and treatment of CDI to address this critical health concern.

In recent years, researchers have made significant efforts in the

development of vaccines against *C. difficile*. Among these, there has been considerable interest in vaccine candidates targeting key surface antigens such as toxins, Cpw66, Cwp84, CD0873, SplA, etc. Studies have shown that these vaccine candidates can elicit specific immune responses against these crucial virulence factors, thereby preventing bacterial colonization, neutralizing bacterial toxins, and alleviating the severity of CDI symptoms [7–10]. However, no vaccine candidate has proven effective enough to warrant approval for preventing CDI. Hence, there is an urgent need to develop novel *C.difficile* vaccines.

Due to the extensive strain diversity and antigenic variability of *C. difficile*, coupled with the challenges associated with in vitro culture, the identification of potential antigens through conventional biochemical, serological, and microbiological approaches is notably time-consuming and labor-intensive. Notably, recent advancements in the fields of immunoinformatics and reverse vaccinology (RV) have proven to be highly effective in facilitating the rapid identification of potential

antigens on a genomic scale [7]. The combination of these approaches has the potential to identify vaccine targets that traditional methods might not be able to detect, thereby expanding the range of potential antigens. Moreover, they eliminate the need for intricate and time-consuming procedures involved in antigen separation and purification, thereby accelerating the screening process, and enhancing overall efficiency [8,9].

In recent years, mRNA-based vaccines have garnered significant attention from researchers because of their rapid production process, capacity to generate robust immune responses without using live pathogens, and adaptability to quickly adapt to emerging viral variants [10, 11]. These vaccines demonstrate significant immunostimulatory effects by activating key innate immune sensing mechanisms, namely Toll-like receptors (TLRs) and RIG-I-like receptors (RLRs), following the delivery of exogenous mRNA into the cells [12]. This activation can stimulate the release of type I interferons and pro-inflammatory cytokines, which in turn prime antigen-presenting cells, such as dendritic cells (DC) and macrophages (MA), to activate a strong adaptive immune response. The transient expression of mRNA also provides safety advantages by eliminating the risks of genomic integration and attendant long-term effects associated with DNA-based vaccines [13]. This ensures precise control over the amount of antigen produced, lowering the chances of immune system overactivation and autoimmune incidents. Furthermore, mRNA vaccines can efficiently produce antigenic proteins within host cells, mimicking natural infection and eliciting potent immune responses with a broad range of antigenic epitopes [14]. Importantly, this process enables the incorporation of post-translational modifications, like glycosylation, which play a vital role in inducing an optimal immune response [14].

In this study, we employed RV approach in combination with various immunoinformatics tools to design a multi-epitope mRNA vaccine against *C. difficile*. We aim to systematically identify potential antigenic targets within the complete proteomes of 118 *C. difficile* strains and to further predict ideal epitopes for these targets, thereby designing vaccines capable of providing protective immunity against CDI based on these epitopes. The proposed multi-epitope mRNA vaccine holds great promise for inducing protective immunity against CDI through several mechanisms. Firstly, it includes multiple epitopes from different antigenic targets, which broadens and specifies the immune response induced by the vaccine, aiding in the recognition and neutralization of an array of virulence factors of different *C. difficile* strains. Second, the transient nature of mRNA expression ensures safety by avoiding genomic integration and long-term negative effects associated with DNA-based vaccines. Simultaneously, it allows the efficient production of antigenic proteins in host cells, mimicking natural infection and inducing strong immune responses. Furthermore, the vaccine includes epitopes that have the capacity to invoke both antibody-mediated and cell-mediated immune responses, ensuring comprehensive immune defense against CDI.

2. Methodology

2.1. Screening conserved antigenic proteins

2.1.1. Obtaining proteome and identifying core proteins

To identify core proteins among all proteomes, we began by retrieving complete proteome sequences of 118 *C. difficile* strains from the NCBI database on April 10, 2023. Subsequently, we utilized version 1.3 of the Bacterial Pan Genome Analysis (BPGA) tool, using the USEARCH algorithm and adopting a screening criterion where the conservation rate of the protein sequence exceeded 90% for the selection of core protein sequences [15]. The default settings for additional parameters were maintained throughout the process of identifying core protein sequences in the dataset.

2.1.2. Identification of non-homologous proteins to humans

In vaccine development, selecting antigens that do not possess homology to human proteins is crucial to avoid the risk of autoimmune responses or tolerance to the introduced antigens. To eliminate potential candidates that may elicit cross-immunity in humans, we first acquired the Homo sapiens reference proteome (taxid: 9606) from the NCBI database and constructed a local database. We then utilized the BLAST v2.12.0+ program to identify core proteins with homology to those within our local database. Proteins exhibiting homology, defined by an E-value threshold of less than 10-4 and a bit score exceeding 100, were omitted from further consideration. All other parameters remained at their default settings.

2.1.3. Identifying essential proteins, virulence proteins, and antibiotic resistance proteins

Essential proteins play an important role in bacterial survival and growth, making them significant targets for vaccine candidates. In our study, we aimed to identify essential proteins of *C.difficile* using core protein alignments with the Database of Essential Genes 10 (DEG 10) via BLAST v2.12.0+ [16]. Core proteins meeting the criteria of an E-value of less than 10-4, a bit score higher than 100, and a sequence identity greater than 30% were considered essential. DEG 10, known for its comprehensive and up-to-date repository of prokaryotic essential proteins, served as a key resource in this process [16]. Furthermore, considering the pivotal role of virulence and antibiotic resistance proteins in pathogen pathogenicity, these proteins were regarded as potential vaccine targets. The BLASTp tool facilitated the identification of such proteins by aligning included essential proteins against the entries in the Comprehensive Antibiotic Resistance Database (CARD) and the Virulence Factor Database (VFDB) [17,18]. Criteria for filtering included an E-value of 10-4 or lower, a bit score greater than 100, and a sequence identity above 30%. We downloaded the sequence data from DEG, VFDB, and CARD on April 10, 2023.

2.1.4. Analysis of subcellular localization (SLC)

Understanding the subcellular localization of proteins is important for identifying therapeutic targets [19]. Proteins located in the bacterial outer membrane, as well as secreted proteins, show considerable promise as vaccine or diagnostic targets due to their accessibility to pharmacological agents. Hence, we used PSORTb v3.0.2 [20], a SCL predictor specifically designed for all prokaryotes, and Cell-PLoc 2.0 [21], a web server for predicting SCL of proteins across various organisms, to identify surface proteins within the categories of virulence proteins or those contributing to antibiotic resistance. Furthermore, we further validated the SCL of all proteins predicted as outer membrane or extracellular proteins by PSORTb v3.0.2 and Cell-PLoc 2.0 servers through the UniProt database. Only those proteins confirmed as outer membrane proteins or secreted proteins by both prediction tools and validated by UniProt were selected for subsequent analysis.

2.1.5. Prediction of biophysical and chemical characteristics

The antigenicity of surface proteins was evaluated employing two bioinformatics servers, VaxiJen v2.0 [22] and ANTIGENpro [23], and proteins with an antigenicity value of ≥ 0.5 were included based on the assessment. Subsequently, the antigenic proteins were submitted to further prediction for toxicity and allergenicity using ToxinPred 2 [24] and AllerTOP v. 2.0 [25] servers, respectively, with proteins identified as either allergenic or toxic excluded from subsequent analysis. Next, the transmembrane helices within the remaining proteins were predicted through the DeepTMHMM server [26], with only proteins having no more than one transmembrane helix being retained. Signal sequences were then identified using the SignalP 6.0 server [27]. Finally, the molecular weight of the chosen proteins was predicted using the ProtParam tool available on the ExPASy server [28]. Only proteins with a molecular weight below 110 kDa were included to enhance experimental feasibility and facilitate protein purification.

2.2. Predicting and screening epitopes

2.2.1. Predicting and screening cytotoxic T lymphocyte (CTL) epitope

The NetCTL 1.2 server [29] was utilized to predict CTL epitopes, selecting only those with a combined score > 0.75 for subsequent antigenicity assessment using the VaxiJen v2.0 server. Antigenic CTL epitopes with an antigenicity score above 0.5 were then analyzed for toxicity and allergenicity using the ToxinPred [30] and AllerTOP v2.0 servers. Those epitopes characterized as neither allergenic nor toxic were advanced for immunogenicity assessment using the Class I Immunogenicity module available at IEDB [31]. Given the essential role of major histocompatibility complex (MHC) molecules in presenting pathogen-derived peptides to T cells, the binding affinities of the non-toxic, non-allergenic and antigenic epitopes to 27 common human leukocyte antigen (HLA) alleles were quantified using the MHC-I binding module from the IEDB [32]. The percentile rank, which is calculated by comparing a peptide's half-maximal inhibitory concentration (IC50) to those obtained from an array of random peptides in the SwissProt database, served as the metric for this evaluation. Epitopes exhibiting a percentile rank of 0.5% or lower, indicating a high affinity for binding to HLA molecules, were identified as potential candidates for vaccine development.

2.2.2. Predicting and screening helper T lymphocytes (HTL) epitope

In this study, the IEDB's MHC II binding tool was employed to identify HTL epitopes and their corresponding HLA-II binding alleles by using the recommended prediction method consensus 2.2 [33,34]. Epitopes with an antigenicity score greater than 0.5 were subsequently screened for toxic and allergenic properties to exclude undesirable candidates. Interferon-gamma (IFN-γ) and interleukin-4 (IL-4) are crucial immunomodulatory factors in the host immune system. IFN-γ is essential for MA activation and the stimulation of natural killer (NK) cells [35], whereas IL-4 is vital for the differentiation of naive T cells into effector cells and for mediating humoral immunity [36]. To evaluate the potential of HTL epitopes to induce the generation of IFN-γ, the analysis was conducted using the IFNepitope server [37]. These inducers were further evaluated for their capacity to prompt IL-4 release by employing the IL4Pred server [38]. Finally, epitopes having the ability to stimulate IFN-γ production were selected for inclusion in the vaccine construct.

2.2.3. Predicting and screening B-cell epitopes

To predict linear B-cell epitopes within the target protein sequences, we used the ABCpred server[[39], setting the epitope length to 16 amino acids with a cutoff value of 0.8. Antigenicity predictions for the selected epitopes were then performed using the VaxiJen v2.0 server. Epitopes achieving an antigenicity score of 1 or higher underwent subsequent analysis for toxicity and allergenicity, employing the Toxinpred and AllerTOP v2.0 servers, respectively. Only those epitopes that were both non-toxic and non-allergenic, with an antigenicity score of 1 or greater, were incorporated into the final vaccine construct.

2.3. Multi-epitope vaccine design

All epitopes that satisfied the established screening criteria were selected for inclusion in the vaccine construct. Due to the relatively weaker immunogenicity of multi-epitope vaccines compared to conventional vaccines, adjuvants are often integrated into vaccine constructs to augment immunogenicity [40]. Beta-defensin-2 demonstrates broad-spectrum antimicrobial activity against bacteria, fungi, and viruses, while also participating in the regulation of inflammation and immune responses. It can activate innate immune cells and recruit naive T cells via chemokine receptors CCR-6 and CCR2 [41]. Moreover, studies have demonstrated its interaction with TLRs on antigen-presenting cells [41]. RS09, a synthetic mimic of lipopolysaccharide (LPS) peptides, serves as a new TLR4 agonist for adjuvant development. When incorporated into certain vaccines as an adjuvant, it

effectively enhances the levels of specific antibodies in the serum [42]. Pam2Cys, a synthetic metabolizable lipoamino acid, mimics the structure of bacterial cell walls, thereby augmenting vaccine immunogenicity through the activation of TLR2 and the stimulation of innate immune responses [43]. Phenol-soluble modulin alpha 4 (PSMα4), produced by bacteria such as Staphylococcus aureus, not only activates the host's immune response and enhances immunity against infections by mimicking biologically active substances during bacterial infections but also triggers the TLR2 signaling pathway by binding to TLR2 [44]. Hence, RS09, Pam2Cys, β-defensin-2, and PSMα4 were selected as adjuvant construct vaccines in this study. The vaccine candidate also included a synthetic peptide called pan HLA DR-binding epitope (PADRE), which acts as a T helper epitope and facilitates the presentation of other epitopes to T cells [45]. EAAAK linker, a rigid peptide linker, efficiently maintains a fixed distance between functional domains, thereby minimizing interference and maintaining individual functional properties [46]. Hence, to maintain individual functional properties of adjuvant and PADRE, we used the EAAAK to connect them. The GPGPG linker can disrupt the formation of junctional epitopes, thereby aiding in the restoration of the immunogenicity of individual epitopes [47]. Additionally, the study by Brian et al. has shown the capacity of GPGPG linkers to provoke immune responses from HTL [48]. The AAY linker, serving as a proteasomal cleavage site in mammalian cells, facilitates the formation of natural epitopes while preventing the formation of junctional epitopes, thereby enhancing epitope presentation and vaccine immunogenicity [47]. The KK linker, composed of lysine residues, provides both flexibility and spatial separation between functional domains or epitopes within protein constructs [49]. This prevents steric hindrance and promotes proteolytic cleavage, facilitating the targeted release of individual components. Therefore, the GPGPG, AAY, and KK linkers were used to link the HTL, CTL, and B-cell epitopes, respectively. Finally, to enable more efficient purification without compromising the vaccine's efficacy, a "6xHis tag" was attached to the fusion protein's C-terminal using the RVRR linker.

2.4. Prediction of the biophysical and chemical properties of the vaccine construct

The antigenicity of vaccine candidates was assessed using VaxiJen v2.0 and validated with the ANTIGENpro server. Additionally, the physicochemical attributes of the vaccine candidates, including molecular weight, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY), were evaluated using the ProtParam tool. Considering that transmembrane domains might impede peptide exposure and complicate peptide preparation, we utilized the DeepTMHMM server to predict transmembrane helices in vaccine candidates employing a deep-learning approach. Furthermore, to prevent signal peptidase cleavage of vaccine sequences, the SignalP-6.0 server was employed to predict signal peptides and their cleavage sites based on the input vaccine sequence. Finally, the solubility of the vaccine was predicted using the SOLpro online server to evaluate its bioavailability.

2.5. Predictions of secondary and three-dimensional (3D) structure, refinement, and validation of 3D structures

The PSIPRED 4.0 server [50] was used for predicting the local conformation of the vaccine backbone, including elements such as alpha helices, beta stand, and coils. The Robetta server was used to model the 3D structures of the vaccine constructs [51]. It employs the homology modeling algorithm when there are structurally similar templates available for the target protein, and the de novo modeling algorithm when such templates are not available [52]. The tertiary structures were subsequently refined using the GalaxyRefine tool on the GalaxyWEB server. This tool iteratively optimizes both the backbone conformation and sidechain orientations, yielding five refined models [53]. The

MolProbity score was used as a reference to screen the best refined 3D model for each vaccine candidate, favoring models with lower scores [54]. Following this, the models underwent quality validation via the ERRAT, and PROCHECK, Verify 3D programs, alongside the ProSA-Web server [55–57]. ERRAT program predicts a score by analyzing the distribution of non-bonded interactions among various atom types within the protein model, serving as an indicator of the 3D structure's quality. A higher ERRAT score indicates a closer match to well-refined protein structures, thus reflecting superior quality [54]. PROCHECK program assesses the stereochemical quality of the 3D models by generating Ramachandran plots, where a higher percentage of residues falling within the favored regions indicates higher quality [54].Verify 3D tool assesses the quality of the 3D structure by comparing the chemical environment of amino acid residues in the model with typical values from a known high-resolution structure database. Models demonstrating higher similarity between residue environments and database typical values indicate higher quality [54]. The overall quality of the 3D structure was evaluated using the Z-score computed by the ProSA-Web web service, where negative values suggest better quality [54].

2.6. Vaccine-toll-like receptors (TLRs) docking

TLR2 and TLR4, found extensively in immune cells, play vital roles in recognizing and responding to specific molecular patterns associated with pathogens, particularly bacteria [58]. To clear the interaction models between vaccine candidates and TLRs (TLR2 and TLR4), molecular docking was conducted using the ClusPro 2.0 server [59]. This server, utilizing a global search algorithm, generated lots of docking models, from which the 10 most favorable were listed based on their minimal energy scores. For enhanced accuracy, the top-ranking docking models from each case underwent further refinement via the refinement module of the HADDOCK 2.4 server [60], which involved optimizing the side-chain conformations of interface residues and performing an energy minimization for the entire complex. Next, interaction residues within the vaccine-TLR complex were predicted via the PDBsum server [61] and visualization of the docking complexes was accomplished with the ChimeraX server [62].

2.7. Analyses of molecular dynamics (MD) simulation and MM-PBSA

2.7.1. MD simulation

MD simulations were performed on docking complexes using Gromacs v2022.1 software to understand intermolecular interactions and dynamic behavior, thus enhancing comprehension of molecular structure, stability, flexibility, and temporal changes over time [63]. Initially, the Amber14SB_parmbsc1 force field parameters were added to the system to generate the topology and coordinate files [64]. The complex systems were then positioned within a cubic simulation box and solvated with transferable intermolecular potential 3P (TIP3P) water molecules. Additionally, Cl-ions were also added to neutralize the overall charge of the system. Next, the energy minimization of the system was performed for 100ps using the steepest descent algorithm. Following that, the NVT (canonical ensemble) and NPT (isobaric-isothermal ensemble) were run 200ps and 2ns, respectively. In the NVT simulation, a modified Berendsen thermostat was employed to gradually increase the temperature from 0 K to 310 K [65], and the NPT simulation utilized the Parinello-Rahman barostat to adjust the pressure to 1 atm [65]. Finally, the 100 ns MD simulation of each complex system was performed under constant temperature and pressure conditions, and the trajectories of the MD simulation were extracted for additional analyses, including the root mean square deviation (RMSD), the radius of gyration (Rg) and hydrogen bonds of each system as well as of root mean square fluctuation (RMSF) of each receptor and ligand chain [66,67]. For Gibbs free energy analysis, Principal Component Analysis (PCA) focusing on PC1 and PC2 was conducted using the gmx covar and gmx anaeig programs, with the resulting principal components utilized in the gmx sham

program to calculate free energy [68,69]. The Gibbs Gibbs free energy landscape (FEL) plots generated using DuIvyTools v0.48 (https://doi.org/10.5281/zenodo.6340263), displayed stable states or energy minima in blue. DCC analysis was carried out using the Bio3D package in R v4.22 software to understand dynamic correlations and cooperative motions between different atoms within the vaccine-receptor complex molecular system.

2.7.2. Molecular Mechanics Poisson-Boltzmann surface area calculation

The Gmx_MMPBSA v1.56 tool was employed to compute the binding free energy of vaccine-TLR complexes using the Molecular Mechanics Poisson-Boltzmann and surface area (MM-PBSA) method [70]. The algorithm used for the calculations is as follows:

$$\Delta G_{bind,slov} = \Delta G_{bind,vacuum} + \Delta G_{slov,complex} - (\Delta G_{slov,vaccine} + \Delta G_{slov,receptor})$$

2.8. Calculation of population coverage

An epitope triggers an immune response only in individuals expressing an MHC molecule able to bind to that specific epitope. The prevalence and distribution frequencies of various MHC alleles, however, differ significantly across regions and ethnic groups. To assess the potential beneficiary population for the designed vaccine, this study employed the Population Coverage program available in IEDB [71]. The analysis incorporated data on the prevalence of HLA genotypes in diverse geographical regions, along with information on MHC-binding capabilities and T-cell restriction [71].

2.9. Immune simulation

The immunogenicity profile of the designed vaccine was predicted using the C-ImmSim server, which employs stochastic differential equations to simulate various immune system components and their interactions with peptides [72]. The C-ImmSim takes into account a variety of factors, such as antigen presentation efficiency, immune cell proliferation, and TH1/TH2 cytokine levels, to assess the immunogenicity potential of the submitted antigen. According to previous studies, the simulation steps were set to 1050 (1 step corresponds to 8 h), and the time step of injection was set at 1, 84, and 168, respectively, allowing for the evaluation of the immune dynamics at different time points [73,74]. All other simulation parameters were kept at the default settings.

2.10. Construction of mRNA vaccine and secondary structure prediction

The design of the mRNA vaccine incorporated several critical elements. First, the Tissue Plasminogen Activator (tPA) in conjunction with the MHC I-targeting domain (MITD) sequence were added to the N-terminal of the WLcd1 protein sequence employing the EAAAK linker. The tPA serves as a guiding element for the sequence to traverse the cell membrane, while the MITD aids in the presentation of MHC molecules. These fused sequences underwent codon optimization based on the codon usage of Homo sapiens, a process carried out using the JCat server [75]. The Codon Adaptation Index (CAI) and the GC content served as metrics for evaluating expression levels. Next, we incorporated the Kozak sequence for translation initiation, along with the TAA stop codon. Furthermore, we added the 5′- untranslated regions and 3′- untranslated regions to improve the mRNA vaccine's stability. Finally, the entire DNA sequence was converted into mRNA utilizing the Transcription Tool, and this was followed by the prediction of the RNA secondary structure with the utilization of the RNAfold server [76].

2.11. In silico codon cloning

The mRNA vaccine was inserted into the pET28a (+) plasmid using SnapGene software, and BamHI and XhoI restriction sites were added to
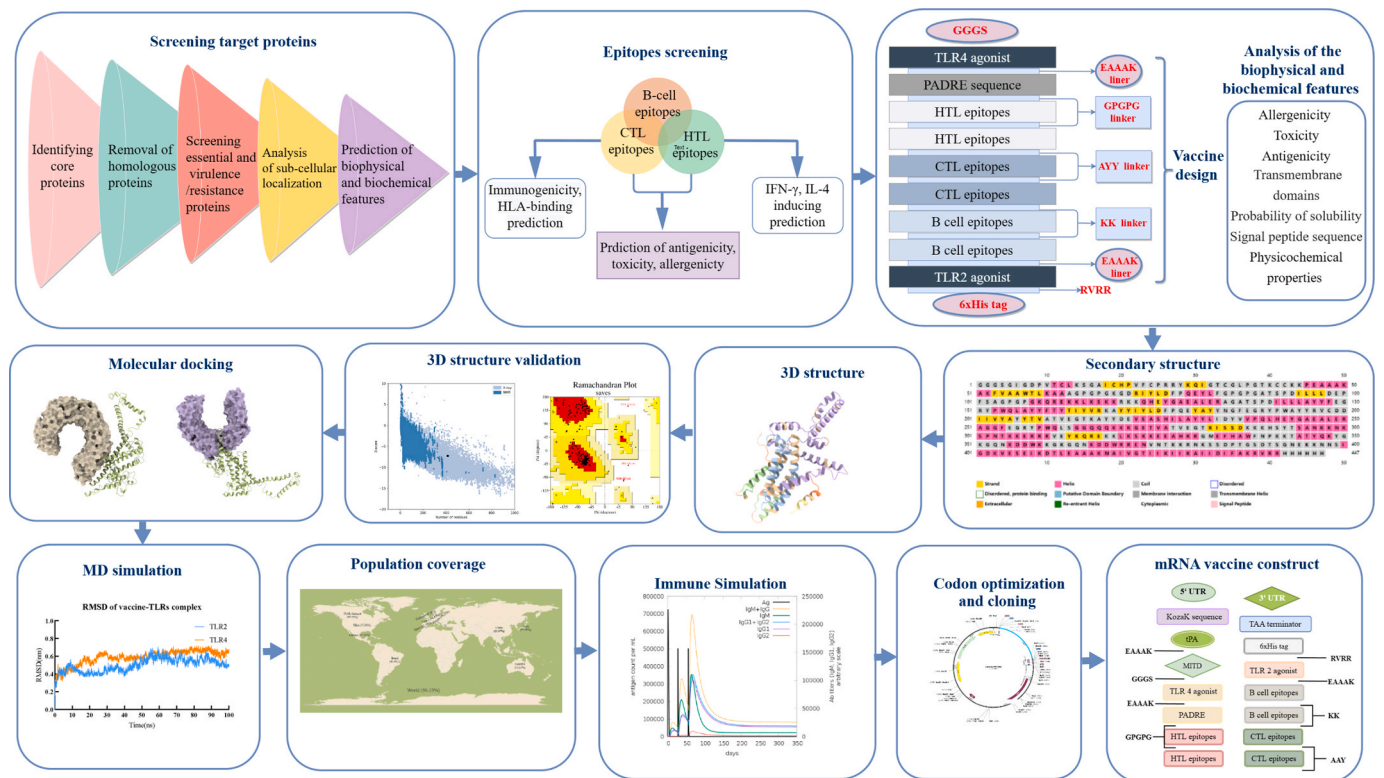
Fig. 1. Flowchart outlining the design and evaluation process for the vaccine construct.

Table 1
Properties of four selected target proteins.

Protein ID	Protein name	SCL[a]	SCL[b]	Antigenicity[a]	Antigenicity[b]	Allergenicity	Toxicity	Transmembrane helix	Location of Signal peptide	MW
Q183K2	Cell surface protein cwp19	Cell wall	Cell wall	0.5234	0.946981	NA	NT	0	24–25	77746.71
Q18A62	ABC-type transport system	Cytoplasmic Membrane	Cell membrane	0.5201	0.53144	NA	NT	0	0	41510.96
Q181X7	peptidoglycan glycosyltransferase	Cytoplasmic Membrane	Cell membrane	0.6085	0.666607	NA	NT	1	0	38960.96
Q180Z8	peptidylprolyl isomerase	Cytoplasmic Membrane	Cell membrane	0.5279	0.905189	NA	NT	0	19–20	37958.21

Protein ID: ID of the protein in the Uniprot; SCL[a]: Sub-cellular localization of target proteins predicted by PSORTb v3.0.2 server; SCL[b]: Sub-cellular localization of target proteins predicted by Cell-PLoc 2.0 server; Antigenicity[a]: Antigenicity predicted by VaxiJen v2.0 server; Antigenicity[b]: Antigenicity predicted by ANTIGENpro server; NT: Non-toxic; NA: Non-allergenic; MW: Molecular weight.

the N-terminal and C-terminal, respectively. The mRNA vaccine sequence was colored in black, while the pET28a (+) vector was displayed in blue. Fig. 1 illustrated a flowchart outlining the vaccine design process and characteristics assessment.

3. Results

3.1. Collection of proteome data and selection of target proteins

A total of 118 complete proteomes of the *C. difficile* strain were obtained from the NCBI database, and their detailed information, including the organism name, accession number, gene count, etc. were presented in Table S1. Within these proteomes, 2149 core proteins were identified (Table S2). After removing proteins homologous to human proteomes, 1834 proteins were remained (Table S3). BLASTp analysis revealed that 582 core proteins were essential for the bacteria's growth and survival. (Table S4). Additionally, 130 proteins were identified as virulence proteins (Table S5), 82 as resistance proteins (Table S6), and

28 exhibited both virulence and resistance properties. After evaluating factors such as SLC, antigenicity, transmembrane domains, signal sequence, allergenicity, toxicity, and physicochemical properties, four optimal proteins were finally selected for vaccine design, as detailed in Table 1. The detailed screening process was depicted in Fig. S1.

3.2. Epitope selection

Using the NetCTL 1.2 server, we identified 351 CTL epitopes with a combined score of > 0.75 from four targeted proteins. Each epitope underwent evaluation for antigenicity, immunogenicity, toxicity, and allergenicity, 40 ideal CTL epitopes were shortlisted. Following that, the affinity of CTL epitopes to HLA-I alleles was assessed. Finally, 9 CTL epitopes were chosen for vaccine design due to their high antigenicity and immunogenicity, absence of toxicity and allergenicity, and strong affinity to class I HLA alleles (Table 2, Table S7).

The MHC II binding tool predicted a total of 1497 HTL epitopes of four target proteins. Among these, 114 HTL epitopes with the Rank≤2%

Table 2
Properties of final selected CTL epitopes.

Protein ID	Start position	Supertype	Epitope	Antigenicity	Immunogenicity	Toxicity	Allergenicity	Rank%	Allele
Q18A62	148	A1, A2	ATSPDILLL	1.3075	0.06618	NT	NA	0.16	HLA-B*58:01
								0.18	HLA-A*02:06
								0.26	HLA-A*32:01
								0.35	HLA-B*57:01
								0.41	HLA-A*68:02
	125	B39, B44	FEGRYPWQL	1.5047	0.1683	NT	NA	0.12	HLA-B*40:01
	303	A3	FTYTIYVRK	0.9696	0.22906	NT	NA	0.05	HLA-A*11:01
								0.06	HLA-A*68:01
								0.08	HLA-A*03:01
								0.29	HLA-A*30:01
	338	A24	IYLDFPQEY	0.6985	0.08302	NT	NA	0.03	HLA-A*30:02
								0.04	HLA-A*23:01
								0.05	HLA-A*24:02
								0.41	HLA-B*35:01
								0.45	HLA-A*32:01
	123	B58	NGFEGRYPW	1.4402	0.21087	NT	NA	0.28	HLA-B*53:01
								0.42	HLA-B*58:01
	198	A1,A3,A26,B62	RVCDDIIVY	1.1737	0.28814	NT	NA	0.05	HLA-A*30:02
								0.08	HLA-B*15:01
								0.14	HLA-A*32:01
								0.34	HLA-B*58:01
								0.37	HLA-B*35:01
								0.41	HLA-A*11:01
								0.43	HLA-A*01:01
Q180Z8	178	B39, B44	DEVEASHIL	0.7833	0.10158	0.10158	NA	0.14	HLA-B*40:01
								0.21	HLA-B*44:03
								0.32	HLA-B*44:02
	27	A3	TVATVEGTK	0.8731	0.23749	0.23749	NA	0.15	HLA-A*68:01
								0.3	HLA-A*11:01
Q183K2	288	A2	LIDYVVPQL	0.5718	0.0055	0.0055	NA	0.1	HLA-A*02:06
								0.12	HLA-A*02:01

Protein ID: ID of the protein in the UniProt; The start position indicated the beginning of the epitope's first amino acid residue; Epitopes with the Rank% ≤0.5 % were considered to have good binding ability to corresponding alleles; NT: Non-toxic; NA: Non-allergenic.

Table 3
Properties of final selected HTL epitopes.

Protein ID	Start position	peptide	Antigenicity	Toxicity	Allergenicity	IFN-γ	IL-4	Adjusted rank	allele
Q18A62	334	KGDRIYLDFPQEYLF	0.5153	NT	NA	IFN-γ inducer	IL4 inducer	0.66	HLA-DRB1*03:01
			0.5153					1.4	HLA-DRB3*01:01
Q18A62	148	ATSPDILLLDEPFSA	0.7975	NT	NA	IFN-γ inducer	Non IL4 inducer	1.6	HLA-DRB1*03:01
Q181X7	89	KQREKKLKSKKRKKQ	1.5953	NT	NA	IFN-γ inducer	IL4 inducer	0.97	HLA-DRB5*01:01

Protein ID: ID of the protein in the UniProt; The start position indicated the beginning of the epitope's first amino acid residue; Epitopes with the Rank% ≤2% were considered to have good binding ability to corresponding alleles; NT: Non-toxic; NA: Non-allergenic.

Table 4
Properties of final selected B-cell epitopes.

Protein ID	Sequence	Start position	Score	Antigenicity	Toxicity	Allergenicity
Q18A62	GFEGRYPWQLSGGQQQ	124	0.87	1.4370	NT	NA
Q180Z8	KGETVATVEGTKISSD	24	0.89	1.1192	NT	NA
Q181X7	KHSYTSANKKNKSPNT	28	0.87	1.1701	NT	NA
	ERRRVEYKQREKKLKS	82	0.81	1.1380	NT	NA
Q183K7	EEAHKRGMEFHAWFNP	115	0.91	1.2577	NT	NA
	TATYQKYGKGQNKDDW	207	0.91	1.0504	NT	NA
	GKGQNKDDWRRENVNT	214	0.89	1.4394	NT	NA
	RNKSSDPTGSDTSGNE	256	0.85	2.1752	NT	NA
	NNSIGDKVESEIKDTL	483	0.84	1.1551	NT	NA

Protein ID: ID of the protein in the UniProt. The start position indicated the beginning of the epitope's first amino acid residue; NT: Non-toxic; NA: Non-allergenic.

were selected for further evaluation. 3 HTL epitopes were finally selected based on specific inclusion criteria, including antigenicity, non-toxicity, non-allergenicity, and the ability to induce IFN-γ production (Table 3, Table S8).

ABCpred server identified a total of 70 B cell epitopes with a high predicted score (≥0.8). After the assessment of antigenicity (≥1), toxicity, and allergenicity, 9 B cell epitopes were finally chosen (Table 4, Table S9).

3.3. Biophysical and biochemical profiles of vaccine candidates

3.3.1. Assessment of the antigenicity, allergenicity and toxicity

This study has developed two vaccine candidates, WLcd1 and WLcd2, that have the same components, except for the different adjuvants at the N- and C-terminals (Fig S2A, S3A). The antigenicity of these two vaccine candidates was highly close. The VaxiJen and ANTIGENpro servers predicted the antigenicity of WLcd1 to be 1.09 and 0.87,
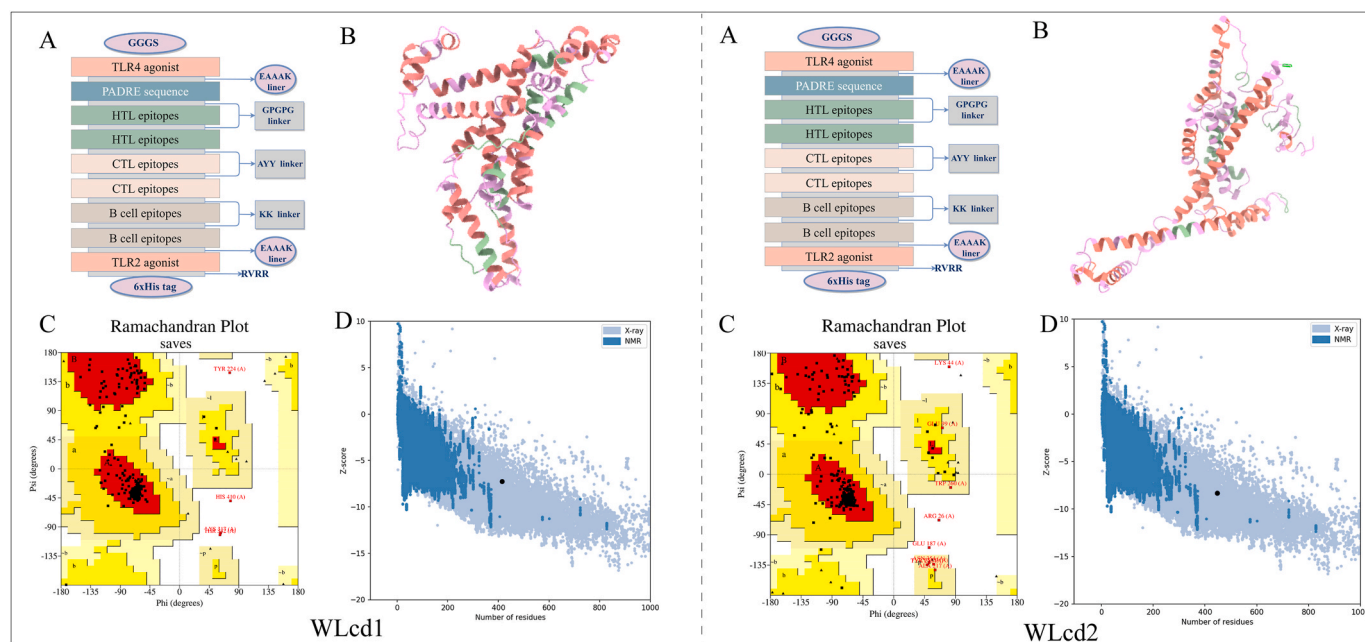
Fig. 2. Design, tertiary structure refinement, and validation of two vaccine constructs (WLcd1 and WLcd2). (A) Design of two vaccine constructs illustrating CTL (cytotoxic T lymphocyte), HTL (helper T lymphocyte), B-cell epitopes, adjuvant, and various linkers, each depicted using different colors. (B) Refined tertiary structures of the two vaccine constructs, with helices depicted in orange-red, strands in green, and coils in purple. (C) Residue distribution in Ramachandran plots generated by the PROCHECK program. WLcd1: 93.6% in favored region (red), 5.3% in allowed region (yellow), and 1.1% in disallowed region (white). WLcd2: 90.8% in favored region (red), 8.1 % in allowed region (yellow), 1% in disallowed region (white); (D) Z-Score plot generated by the ProSA-Web program, indicating Z-scores of −7.27 and −7.46 for WLcd1 and WLcd2, respectively.

respectively. Similarly, WLcd2 showed antigenicity values of 1.02 and 0.88, respectively (Table S10). Based on the results obtained from the ToxinPred2 and AllerTOP v.2.0 servers (Table S10), it was determined that both the WLcd1 and WLcd2 were non-toxic and non-allergenic.

3.3.2. Analysis of physicochemical characteristics

The vaccine candidates WLcd1 and WLcd2 had residue lengths of 414 and 447, respectively. Their molecular weights were 47,353.17 Da and 50,688.58 Da, respectively, both falling within the recommended weight range (≤110 KDa) (Table S10). The instability index values for WLcd1 and WLcd2 were 34.71 and 34.95, respectively. A protein with an instability score of 40 or higher indicates the potential of being unstable and susceptible to degradation. The aliphatic index values for WLcd1 and WLcd2 were 55 and 59.89, respectively. The aliphatic index reflects thermal stability, with favorable values typically ranging between 50 and 150. The GRAVY index for WLcd1 was −1.011, and for WLcd2 was −0.876, indicating hydrophilicity for two vaccine candidates (Table S10). Predictions from DeepTMHMM and SignalP-6.0 showed the absence of transmembrane helices or signal peptides in either vaccine (Fig. S4-5). The solubility propensity for WLcd1 and WLcd2 upon overexpression in *E. coli* was 0.965418 and 0.907515, respectively (Table S10). In summary, our findings indicated that both vaccines possess favorable biophysical and biochemical attributes suitable for immunological applications.

3.4. Prediction of secondary Secondary structure

The PSIPRED server was employed to analyze the secondary structure of WLcd1 and WLcd2. The analysis revealed that 41.06% of the amino acids in WLcd1 contributed to alpha-helix formation, 43.72% to coils, and 15.22% to beta-strand. Similarly, for WLcd2, 43.85% of its amino acids formed alpha-helices, 43.62% were involved in coil formation, and 12.53% contributed to beta-strand formation (Fig. S2B, 3B).

Table 5
The docking analysis of the vaccine-TLR complexes.

Server	Parameter	Complex			
		WLcd1-TLR2	WLcd1-TLR4	WLcd2-TLR2	WLcd2-TLR4
ClusPro	Center	−890.8	−855	−903.0	−853.2
	Lowest energy	−1041.5	−1331.7	−1118.6	−1042.6
HADDOCK 2.4	HADDOCK score	−247.0 ± 5.8	−222.4 ± 1.8	−203.8 ± 2.7	−203.0 ± 3.6
	Cluster size	20	20	20	20
	RMSD	0.6 ± 0.3	0.5 ± 0.3	0.6 ± 0.3	0.6 ± 0.3
	Van der Waals energy	−113.3 ± 2.9	−87.5 ± 2.2	−100.8 ± 3.1	−96.9 ± 3.0
	Electrostatic energy	−232.1 ± 27.4	−435.2 ± 8.2	−207.6 ± 10.5	−250.1 ± 22.8
	Desolvation energy	−87.3 ± 7.7	−47.8 ± 2.4	−61.5 ± 1.3	−355.3 ± 18.9
	Restraints violation energy	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	Buried Surface Area	3095.1 ± 69.9	3390.8 ± 54.1	2455.9 ± 104.3	3572.4 ± 46.5
	Z-Score	0	0	0	0
PDBsum	Salt bridges	3	4	2	4
	Hydrogen bonds	15	17	6	13

3.5. 3D structure modeling, refinement and validation

The 3D structures of the designed vaccines were generated using the Robetta server, and the most favorable models were selected for refinement via the GalaxyRefine server. After refinement, the models of WLcd1 and WLcd2 exhibited enhancements in quality. The most ideal refined models of WLcd1 and WLcd2 were shown in Fig. 2B. The ERRAT scores of the refined models of WLcd1 and WLcd2 were 96.9697 and 94.8718, respectively. Additionally, the Ramachandran plot of the refined model of WLcd1 showed that 93.6% of residues were in favored
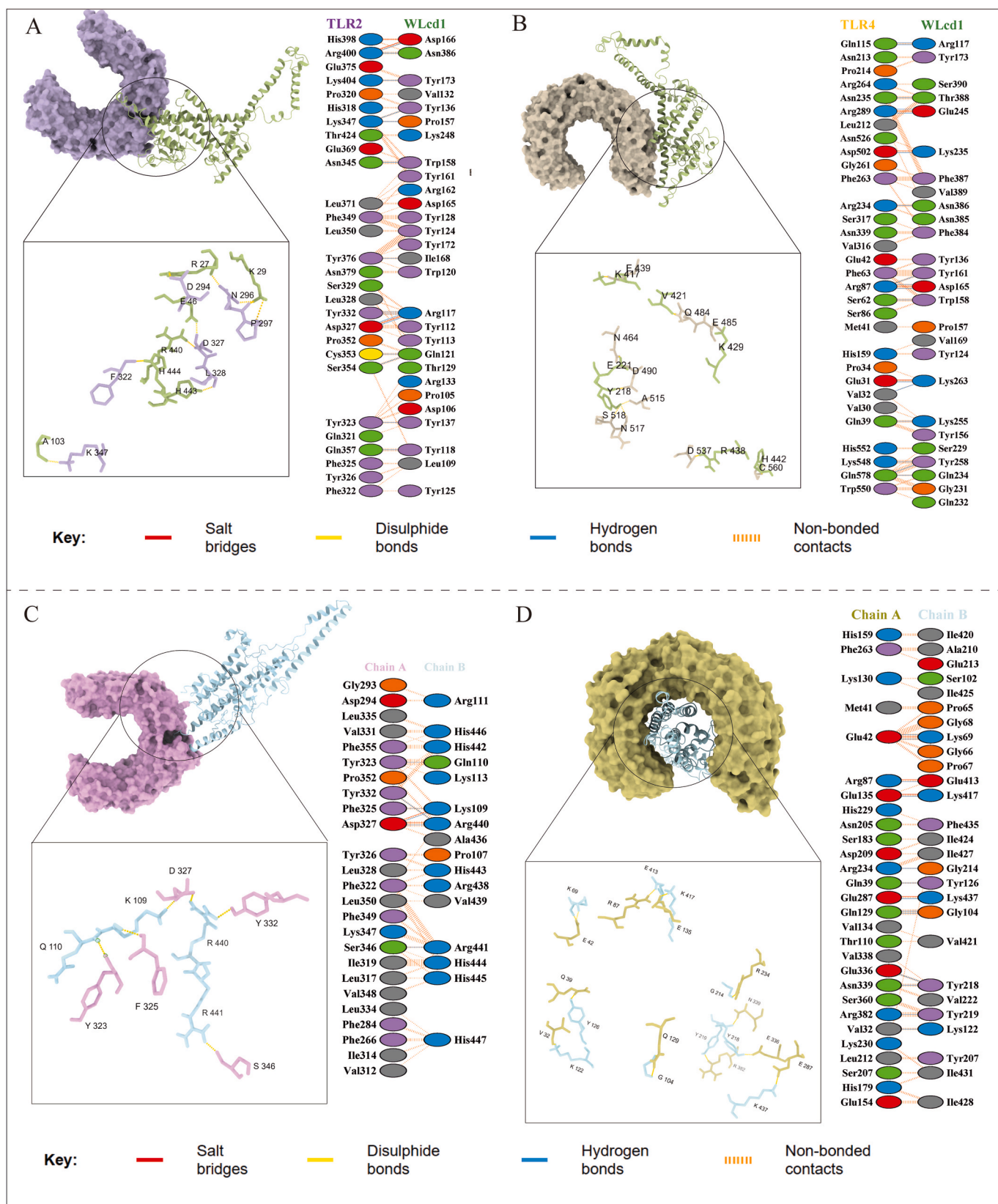
Fig. 3. Molecular docking between vaccine candidates and receptor chains. Docking patterns of the vaccine candidate and receptor were illustrated at the top, with the vaccine chain depicted in cartoon form and the receptor chain displayed in surface form. Hydrogen bonds, shown in the yellow dotted line, formed between the vaccine chain and the receptor chain were presented below. Different interactions formed between them were shown in the plot on the right, with different types of interactions represented by differently colored lines. (A) WLcd1-TLR2 complex (15 hydrogen bonds, 3 salt bridge, and 265non-bonded contacts); (B) WLcd1-TLR4 complex (17 hydrogen bonds, 4 salt bridge, and 157 non-bonded contacts); (C) WLcd2-TLR2 complex (2 hydrogen bonds, 6 salt bridge, and 130 non-bonded contacts); (D) WLcd2-TLR4 complex (4 hydrogen bonds, 13 salt bridge and 119, non-bonded contacts).
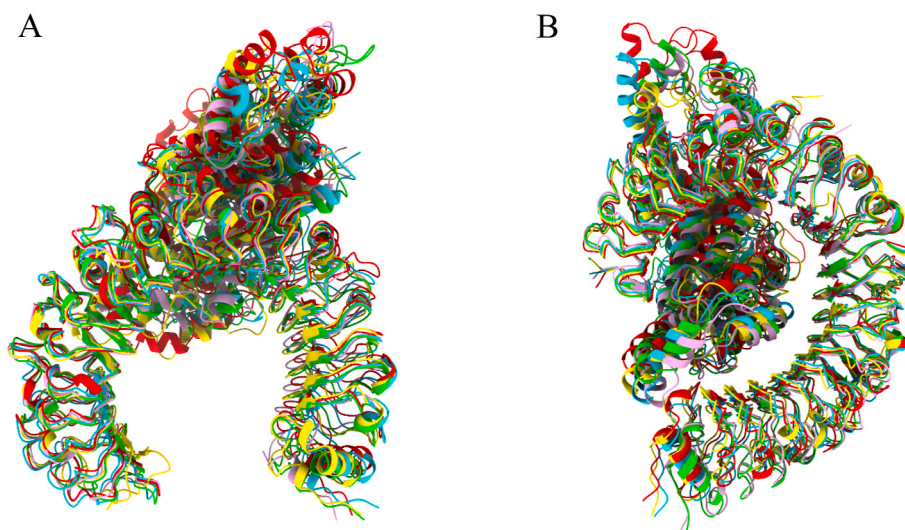
A   B

Fig. 4. Superposition diagram of vaccine-receptor complexes at different time points in molecular dynamics (MD) simulations. The structure diagrams of 0ns, 25ns, 50ns, 75ns, and 100 ns times were shown in red, yellow, blue, green, and purple, respectively. (A) WLcd1-TLR2 complex; (B) WLcd1-TLR4 complex.

regions, 5.30% in allowed regions, and only 1.1% in disallowed regions (Fig. 2C). Similarly, for the refined model of WLcd2, 90.8% of the residues were found in the favored region, 8.1% in allowed regions, and only 1% in disallowed regions (Fig. 2C). Verify3D tool has also verified the reliability of vaccine constructs by demonstrating that 89.37% and 80.31% of residues in the refined models of WLcd1 and WLcd2, respectively, align closely with the average score line. The Z-scores predicted by the ProSA-Web server were −7.27 for WLcd1 and -7.46 for WLcd2, respectively (Fig. 2D). These results suggested the reliability of the refined 3D models for both WLcd1 and WLcd2.

3.6. Molecular docking of vaccine with TLRs

Molecular docking was performed to clarify the binding mode between refined models of vaccine candidates and immune receptors. Thirty models were displayed and ranked according to the cluster size generated by the ClusPro 2.0 server, the docking complex with the largest cluster size was selected (Table 5). The WLcd1-TLR2 complex with the largest cluster size exhibited a center binding score of −890.8, with the lowest energy identified at −1041.5. For the WLcd1-TLR4 complex, the center binding score was −780.2, with the lowest energy observed at −1048.1 (Table 5). In the case of the WLcd2-TLR2 complex, the binding score at the center was −903.0, and the lowest energy was found at −1118.6. For the WLcd2-TLR4 complex, the center binding score was −853.2, and the lowest energy was observed at −1042.6 (Table 5). To further improve the structural stability and reliability of the complexes, we refined the docking complexes using the HADDOCK server. The refined docking models of vaccine-TLR complexes were visualized via ChimerX (Fig. 3). For a thorough comprehension of vaccine-receptor interactions, the refined models of the vaccine-TLR complex were additionally analyzed using PDBsum. According to the PDBsum server predictions, there were 15 hydrogen bonds and 3 salt bridges in the WLcd1-TLR2 complex, while the WLcd1-TLR4 complex displayed 17 hydrogen bonds and 4 salt bridges (Table 5, Fig. 3). In comparison, the interactions between WLcd2 and TLRs were found to be weaker. Specifically, the WLcd2-TLR2 and WLcd2-TLR4 complexes had 6 hydrogen bonds and 2 salt bridges, and 13 hydrogen bonds and 4 salt bridges, respectively (Table 5, Fig. 3). Above all, WLcd1 and WLcd2 had a good binding affinity with TLRs.

3.7. MD simulation

To assess the stability, flexibility, and compactness of the vaccine-

TLR complexes, we performed molecular dynamics (MD) simulations on the docked complexes. Subsequently, we analyzed the RMSD, Rg, RMSF, and hydrogen bonds of the simulated trajectories. To know the structural changes over time during MD simulations, we extracted structures of the docking complexes at 0, 25, 50, 75, and 100 ns and visualized their superimposed structures using ChimeraX (Fig. 4).

3.7.1. Analysis of RMSD

The RMSD serves as a quantitative index used to evaluate the level of conformational changes between a simulated structure and its original state. When the RMSD value is low, it typically indicates minimal conformational changes, while a stable RMSD value suggests relative structural stability. In the case of the WLcd1-TLR2 complex, the RMSD value showed a gradual increase in the first 10 ns, then continued to rise over time, reaching a maximum value of 0.6 nm at 55 ns. After 55 ns of MD simulation, the RMSD value displayed reduced fluctuation, suggesting that the WLcd1-TLR2 complex has achieved stable. Similarly, for the WLcd1-TLR4 complex, the RMSD value showed a gradual increase in the first 10 ns, then experienced another increase at 40 ns, and the system reached relative stability at 56 ns of the MD simulation with a RMSD value of 0.6 nm (Fig. 5A).

3.7.2. Analysis of Rg

The Rg value, calculated as the root-mean-square distance of all atoms in a molecule from their center of mass, is used to characterize the compactness of the protein complex. A stable Rg value indicates that the structure of the molecule remains stable and undergoes no significant structural change. As depicted in Fig. 5B, the Rg value for both the WLcd1-TLR2 and WLcd1-TLR4 complexes remained relatively stable throughout the 100 ns MD simulation, with fluctuations limited to a range of 0.3 nm.

3.7.3. Analyses of RMSF

During MD simulation, the RMSF serves as a measure of the positional deviation of individual atoms within a molecule from their average positions. It's employed to assess the structural flexibility of each amino acid residue within the molecule. Generally, regions exhibiting higher RMSF values indicate greater flexibility within the ligand-receptor complex. The RMSF analysis of both the WLcd1-TLR2 and WLcd1-TLR4 complexes showed that the vaccine chain had more changes in RMSF values than the receptor chain. These changes were especially noticeable in residues after 380, which suggests that these residues were more flexible during the MD simulation (Fig. 5C and D).
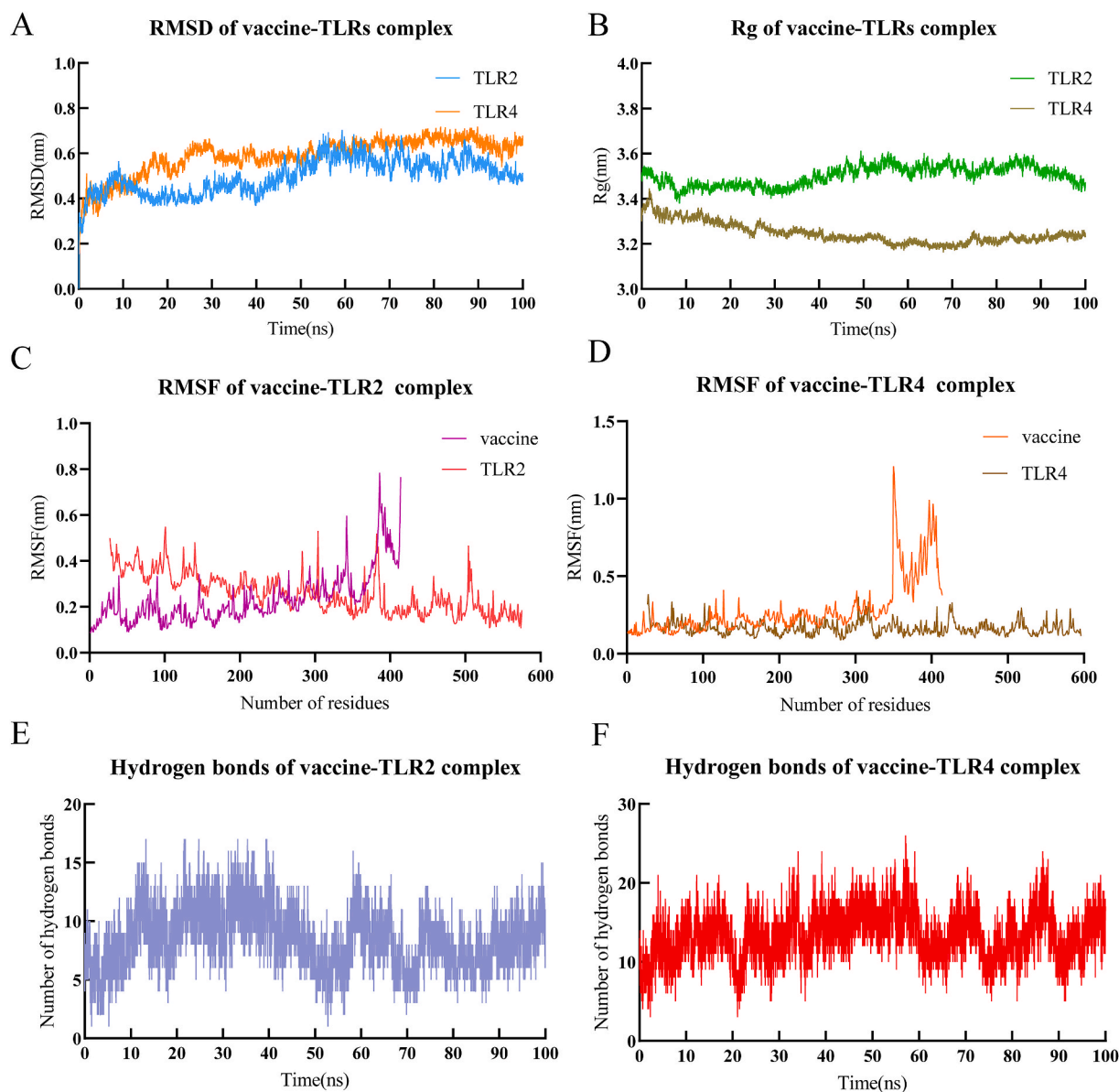
Fig. 5. The Root means square deviation (RMSD), Root means square fluctuation (RMSF), Radius of Gyration (Rg) and hydrogen bonds of molecular dynamics simulation (MD) trajectories of vaccine-TLR complexes. (A)The RMSD of WLcd1-TLR2 (orange) and WLcd1-TLR4 (blue) complexes; (B) The Rg of WLcd1-TLR2 complex (green) and WLcd1-TLR4 complexes (brown); (C) The RMSF of WLcd1-TLR2 complex, the vaccine chain was represented in purple and the TLR2 chain was depicted in red; (D) The RMSF of WLcd1-TLR4 complex, the vaccine chain was represented in orange and the TLR4 chain was depicted in brown; (E) The hydrogen bonds of WLcd1-TLR2 complex; (F) The hydrogen bonds of WLcd1-TLR4 complex.

Table 6
MM-PBSA analysis of the vaccine-TLR complexes.

Energy component	Complex	
	WLcd1-TLR2	WLcd1-TLR4
ΔVDWAALS(KJ/mol)	−130.84(10.29)	−166.64(7.07)
ΔEEL(KJ/mol)	−841.13(74.95)	−2134(81.91)
ΔEPB(KJ/mol)	902.74(71.51)	2189.18(76.65)
ΔENPOLAR(KJ/mol)	−15.37(1.03)	−20.12(0.89)
ΔGGAS(KJ/mol)	−971.97(75.41)	−2300.63(83.03)
ΔGSOLV(KJ/mol)	887.37(71.47)	2169.06(76.37)
ΔTOTAL(KJ/mol)	−84.6(9.08)	−131.58(12.55)

DVDWAALS: van der Waals energy; DEEL: Electrostatic energies; DEPB: Polar solvation energy; DENPOLAR: Nonpolar solvation energy; ΔGGAS = DVDWAALS + DEEL; ΔGSOLV = DEPB + DENPOLAR; ΔDTOTAL = ΔGGAS + ΔGSOLV.

For the TLR2 chain, the RMSF value showed small fluctuations during the MD simulation, with a mean RMSF value of 0.26 ± 0.08 and a maximum RMSF value of 0.5488 (Fig. 5C). The RMSF value for the TLR4 chain remained stable during the MD simulation, with an average value of 0.16 ± 0.04 (Fig. 5D).

3.7.4. Analysis of hydrogen

The hydrogen bond, an intermolecular force, is crucial for stabilizing protein complex structures. During MD simulation, the total number of hydrogen bonds in the WLcd1-TLR2 complex fluctuated around 9, while in the WLcd1-TLR4 complex, it fluctuated around 13. It was evident that interactions within the WLcd1-TLR4 complex were stronger than those in the WLcd1-TLR2 complex. This aligned with our analysis of RMSD values and MM-PBSA binding energy (Table 6). These results indicated that both the WLcd1-TLR2 and WLcd1-TLR4 complexes were stable, and the designed vaccine exhibited a strong affinity to TLRs (Fig. 5E and F).
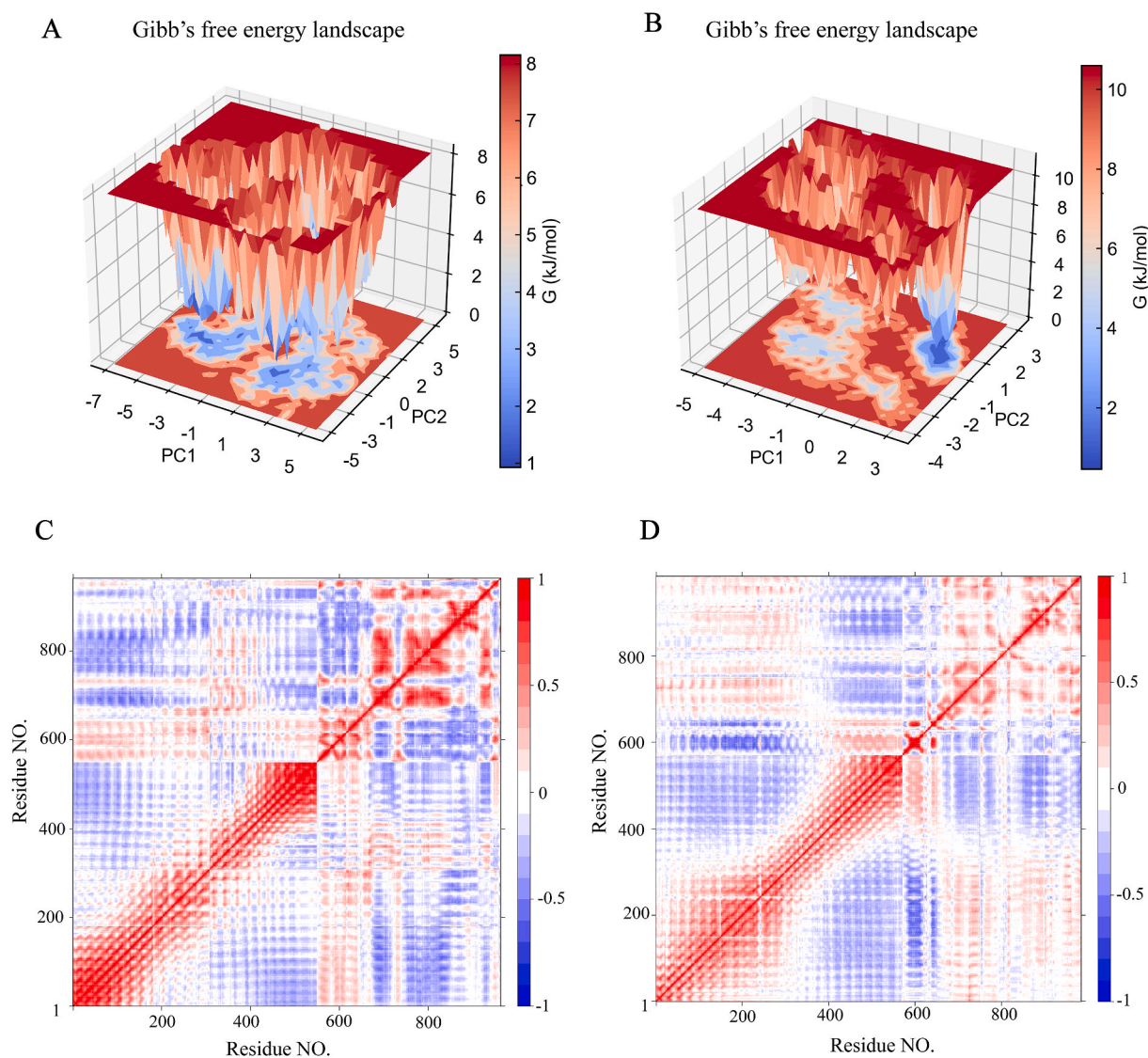
Fig. 6. The plot of Gibb's free energy landscape (Gibb's FEL) and dynamical cross-correlation matrix (DCCM) of vaccine-receptor complexes. The Gibbs FEL plot displays stable states in blue and unstable states in red for vaccine-receptor complexes. In the DCCM plot, red, white, and blue colors represent positive, no, and negative motions, respectively, between paired residues in the two chains. (A) Gibb's FEL of WLcd1-TLR2 complex. (B) Gibb's FEL of WLcd1-TLR4 complex. (C) DCCM of WLcd1-TLR2 complex. (D) DCCM of WLcd1-TLR4 complex.

3.7.5. Gibb's free energy analysis

The Gibbs free energy was employed for analyzing the conformational stability of vaccine-receptor complexes, utilizing the first two eigenvectors (PC1 vs PC2) derived from the MD simulation trajectory. Regions characterized by lower free energy in the Gibbs FEL plot denote stable states, while areas with higher free energy signify transitional states. As illustrated in Fig. 6A, the WLcd1-TLR2 complex system demonstrated a greater number of energy basins at the lowest levels, with the lowest energy conformations concentrated within energy basins ranging from −5.8 to 6.1 kJ/mol on PC1 and -4.6 to 5.1 kJ/mol on PC2. Conversely, the WLcd1-TLR4 system exhibited fewer energy basins at the lowest levels, with the lowest energy conformations spanning from 0.4 to 3.7 kJ/mol on PC1 and -0.6 to 2.3 kJ/mol on PC2 (Fig. 6B).

3.7.6. Dynamic cross-correlation analysis

The DCCM plot (Fig. 6C and D) visually illustrated the dynamic correlations between atoms within vaccine-TLR complex systems, utilizing a color scale to represent both the direction and strength of these correlations. Specifically, in the plot, the color red corresponds to a correlation coefficient of 1, while white represents 0, and blue signifies

−1. As shown in Fig. 6C and D, once the vaccine-receptor complex system reached a stable state, intramolecular movements became predominant throughout the MD simulation, with only sporadic intermolecular movements. As shown in Fig. 6C, parts of the vaccine chain between 0 and 100 had a slightly positive correlation with the TLR2 chain in the WLcd1-TLR2 complex. Similarly, in the WLcd1-TLR4 complex, the vaccine chain's residues 11–114 showed a slightly positive correlation with the TLR2 chain's residues 1–400 (Fig. 6D).

3.7.7. MM-PBSA calculations

We extracted the 90–100 ns MD simulation trajectory at intervals of 100 ps for subsequent MM-PBSA calculations. The calculated total binding free energies for WLcd1-TLR2 and WLcd2-TLR4 were −84.6 kcal/mol and −131.58 kcal/mol, respectively. Detailed energy analyses were presented in Table 6. These results were consistent with the findings from the RMSD and hydrogen bond analyses. The results showed that both WLcd1-TLR2 and WLcd1-TLR4 complexes were stable, with WLcd1-TLR2 being comparatively more stable.
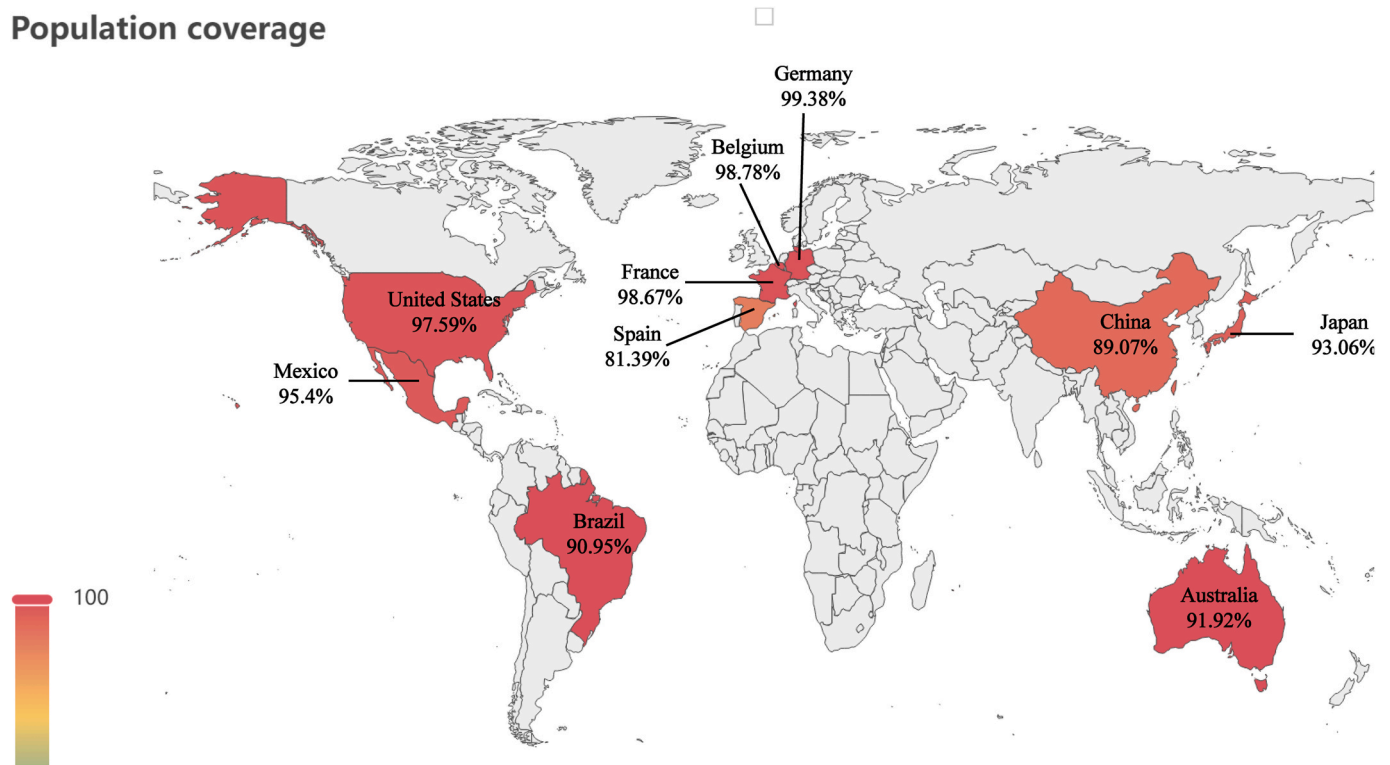
Population coverage

Fig. 7. Population coverage of the vaccine candidate across different areas.

3.8. Population coverage

Population coverage across regions represents the percentage of individuals who would benefit from mass vaccination efforts. According to the predicted results, the designed multi-epitope vaccine (WLcd1) achieved a global coverage rate of 96.53%, with particularly elevated rates observed in North America (97.51%) and Europe (98.51%), regions where *C. difficile* prevalence is higher. We have also calculated the vaccine's population coverage in the countries with a high incidence of CDI (Fig. 7). These findings suggest that the vaccine candidate has the potential to protect a significant portion of the population against CDI, particularly in nations such as the United States and Australia, which experience more severe CDI burdens.

3.9. Immune simulation

To better understand the dynamics of antibody production and cell-mediated immune reactions following vaccination, immune simulations were conducted. These simulations were aimed at mapping the initiation and progression of the immune response within the host organism after vaccination. After administering the first dose, a slight increase in the levels of antibodies (including IgM, IgG + IgM, and IgG1 + IgG2), plasma cells (of the IgM isotype), and B cells (also isotype IgM) were observed, which further amplified after the second and third injections (Fig. 8A–C). 20 days after the administration of the third dose, the concentrations of IgM + IgG and IgG1 + IgG2 reached their peak, estimated at approximately 2.3×10^5 and 1.1×10^5, respectively. Meanwhile, the aggregate populations of B cells and IgM isotype plasma cells peaked at approximately 800 and 90 cells per mm3. Additionally, the elevation of B cells (isotype: lgG1) and plasma cells (isotype: lgG1) was also detected during the second and third immune responses (Fig. 8A–C). In addition to humoral immune response, the activation of cell-mediated immunity was also evident, as demonstrated by an increase in the populations of active cytotoxic T-cells and helper T-cells (Fig. 8D and E). As shown in Fig. 8D, the population of active cytotoxic

T-lymphocytes increased steadily and reached 1000 cells/mm3 around 30 days after vaccination. The growth pattern of active helper T-cells showed a three-stage cascade reaction, peaking at 8000 cells/mm3 around day 50 (Fig. 8E). The generation of cytokines, including IFN-γ, IL-2, and others, was observed following vaccine administration, as indicated in Fig. 8F. Additionally, elevated levels of NK cells, MA, and DC were also noted during the immune simulation (Fig. 8G–I). In summary, WLcd1 appears to be capable of eliciting both adaptive and memory immune responses within the host, in theory.

3.10. Design of the mRNA vaccine and recombinant plasmid

The WLcd1-mRNA vaccine, which included both the tPA and MITD sequences, had a length of 2391 nucleotides. The optimized WLcd1-mRNA vaccine sequences yielded a CAI value of 0.97 and a GC content of 48.31%, both of which are within the desired parameters. The final mRNA vaccine construct was depicted in Fig. 9A. Additionally, the best and central secondary structures of the mRNA vaccine were presented in Fig. 9B and C, respectively, exhibiting minimum free energy of −858.60 kcal/mol and −635.50 kcal/mol, respectively. The recombinant WLcd1-mRNA-pET-28a (+) vector was shown in Fig. S6.

4. Discussion

C.difficile exhibits inherent resistance to a multitude of antibiotics, greatly limiting available treatment options. This challenge is worsened by the spread of highly virulent strains (BI/NAPI/027) and the significant recurrence rate observed after treatment with sensitive antibiotics [2,77]. Recognizing the severity of this issue, the Centers for Disease Control and Prevention (CDC) classified *C. difficile* as an "urgent antibiotic-resistant threat" in 2019 [78]. Active vaccination is widely acknowledged as an effective approach for preventing CDI, yet no vaccine has been approved. The failure of a toxoid vaccine in phase III trials to prevent primary CDI highlighted the necessity for a new vaccine effective in both healthcare and community settings [79]. The
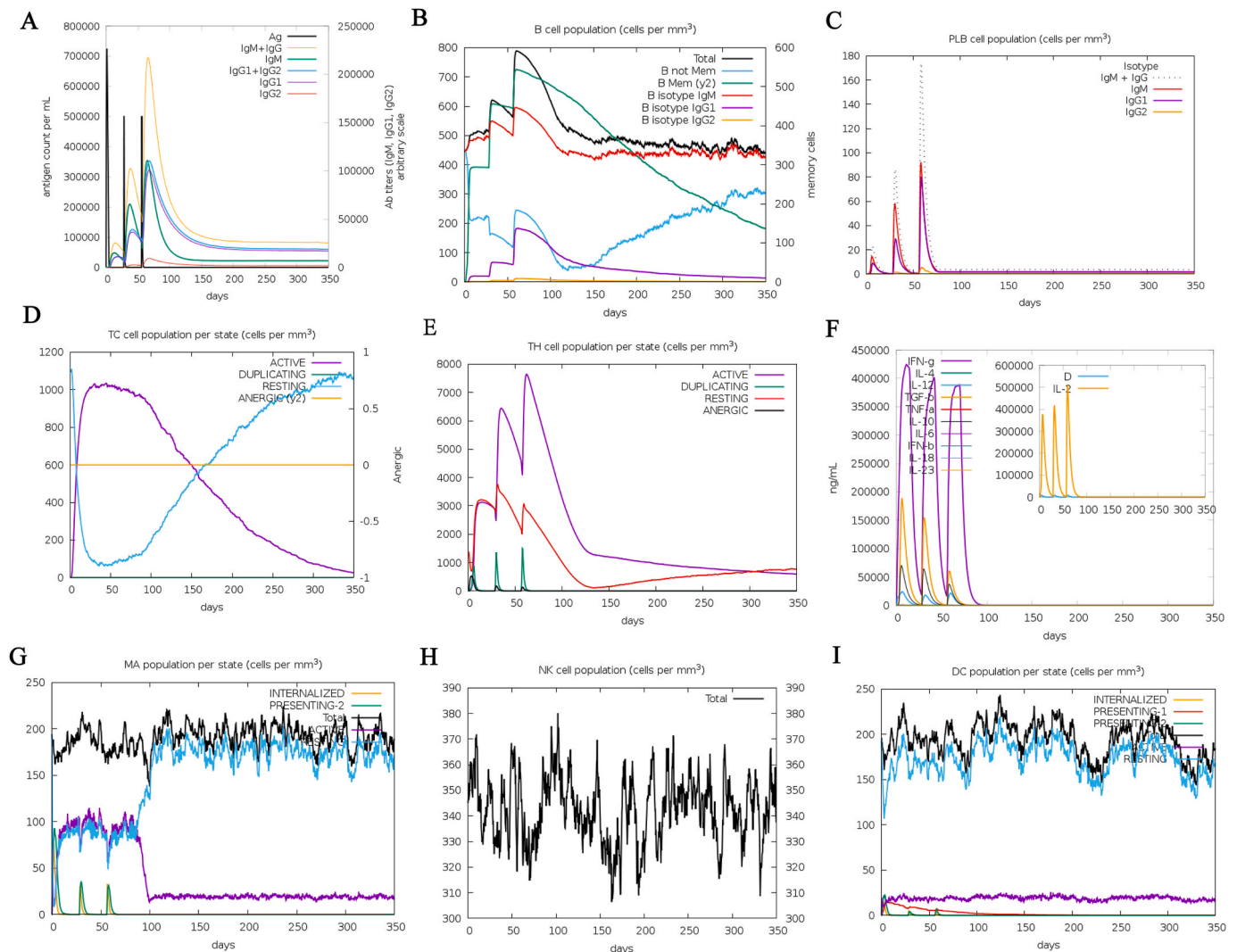
Fig. 8. Immune response simulation elicited by three injections of the vaccine construct (WLcd1) generated by C-ImmSim server. (A) Titers of different antibodies. (B) Levels of B cell population. (C) Levels of plasma B cells. (D) Levels of helper T-cell population. (E) Levels of cytotoxic T cell population. (F) Concentration of cytokines. (G) Levels of NK cell population. (H) Levels of MA population. (I) Levels of DC population.

developmental strategy of a multi-epitope mRNA vaccine facilitates the immune system's recognition of various antigenic determinants, potentially broadening and strengthening the immune defense against *C. difficile*. Moreover, mRNA vaccines can be translated directly into antigen proteins within cells through cellular processes. These newly synthesized antigen proteins are then quickly presented to antigen-presenting cells in the immune system [12]. Hence, the multi-epitope mRNA vaccine may be an ideal alternative to prevent *C. difficile*.

Most previous studies only focused on reference strains like CD630 or epidemic strains like R20291, neglecting the diversity of *C.difficile* strains and their genomes [80,81]. This study diverges by analyzing 118 complete proteome sequences of *C.difficile*, aiming to identify core conserved targets across the different *C.difficile* strains. This approach offered new insights for the development of vaccines capable of broadly preventing CDI. Employing a subtraction proteomics screening approach, we rigorously selected proteins that exhibited high sequence similarity across all selected strains, ensuring their widespread presence and similarity across different strains. Subsequent filtration processes considered various factors, including indispensability, virulence, and resistance, while excluding proteins homologous to human proteins to avoid potential autoimmune risks. Ultimately, we identified four vaccine targets located in the bacterial outer membrane or extracellular

space, characterized by their essentiality, antigenicity, non-allergenicity, and non-toxicity. This approach not only prioritized proteins crucial for bacterial survival and pathogenicity but also minimized the risk of adverse immune reactions post-vaccination. Importantly, all four identified proteins were associated with virulence or resistance functions, highlighting their potential importance in stimulating a protective immune response against *C. difficile*.

The primary goal of the specialized, active acquired immune responses is to eradicate or control the spread of infectious diseases. Memory B cells, generated through adaptive immunity, can recognize specific pathogens upon subsequent contacts, forming the foundation of immunological memory crucial for vaccination efforts [82]. B- and T-lymphocyte cells collaborate within the adaptive immune system to establish antibody-dependent cellular immunity against foreign invaders, ensuring the efficient recognition and elimination of pathogens. Therefore, to ensure the induction of a strong immune response in the host, the designed vaccine incorporated not only T-cell epitopes but also B-cell epitopes. The selection of epitopes for the vaccine candidate was based on their antigenicity, immunogenicity, allergenicity, and toxicity. Additionally, in predicting CTL and HTL epitopes, their affinity to MHC molecules was also assessed, which is consistent with many studies [83, 84]. In vaccine development, incorporating linkers helps preserve the structural stability of the vaccine, enhancing immunogenicity, and
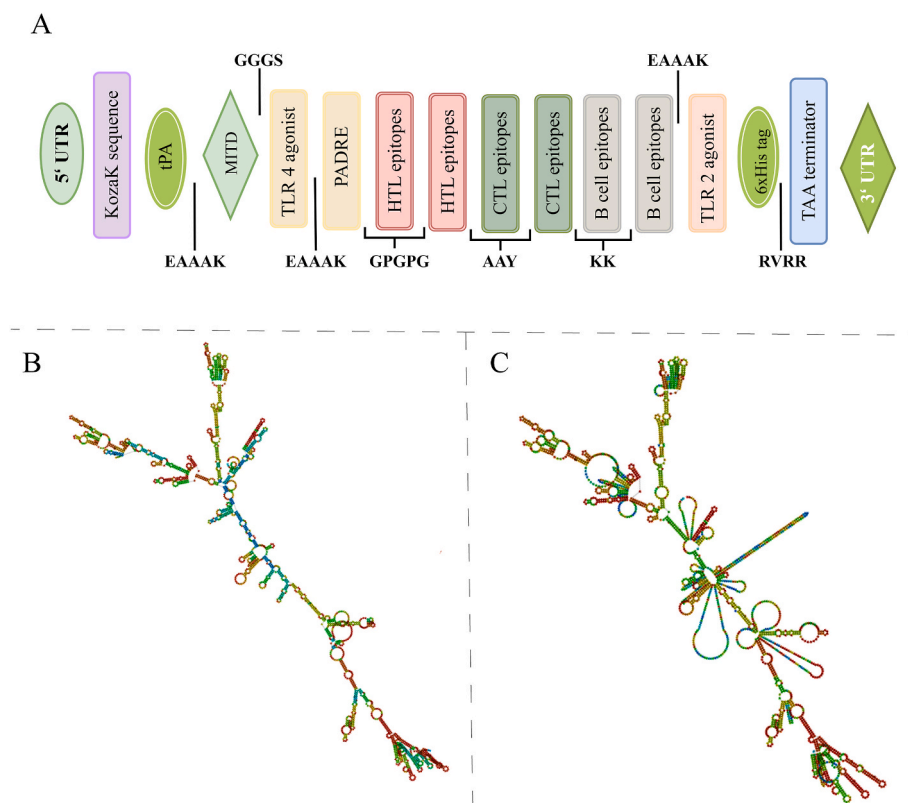
Fig. 9. Design and prediction of mRNA structure of vaccine construct. (A) Design of the mRNA vaccine construct; (B)The best secondary structure of the mRNA vaccine and (C) the centroid secondary structure of the mRNA vaccine generated using RNAfold server.

facilitating proper folding. Hence, EAAAK, AAY, GPGPG, and KK linkers were utilized to connect different parts of the vaccine constructs, as reported in many previous studies [85,86]. To strengthen immune activation and maintain the durability of the immune activation [40], RS09, Pam2Cys, β-defensin-2, and PSMα4, as adjuvants, were used to construct *C.difficile* vaccines, all of which have been utilized in several previous studies [83,87–89].

To ensure the successful implementation of both in vitro and in vivo experiments involving the *C. difficile* vaccine candidates, we conducted a comprehensive assessment of their physicochemical characteristics. The high antigenicity scores and non-toxic, non-allergenic profiles observed in WLcd1 and WLcd2, align with predictions of multiple multi-epitope vaccines that have been experimentally validated [87,88,90]. This consistency further confirmed their safety and potential to induce immune responses. As the vaccine will be administered into the aqueous environment within the host body, achieving high solubility is crucial for its development. Both vaccines' GRAVY and solubility further support their suitability for future applications.

A comprehensive understanding of the molecular interactions between vaccine candidates and immune receptors is crucial for the successful development of vaccines. The molecular docking results demonstrated that refined models of vaccine candidates (WLcd1 and WLcd2) exhibit a strong affinity for TLRs, which are important for inducing immune responses. Notably, the WLcd1-TLR2 and WLcd1-TLR4 complexes had a higher HADDOCK score and buried surface area, indicating that WLcd1 has the potential to be an effective vaccine candidate against *C. difficile*. Although interactions between WLcd2 and TLRs were observed, they appeared to be relatively weaker compared to WLcd1, suggesting differences in the binding affinities of vaccine-receptor complexes. This observation underscores the significance of the structural and conformational characteristics of vaccine candidates in determining their immunogenicity [91,92]. Furthermore, the refinement of docking complexes using the HADDOCK server enhanced their

structural stability, as evidenced by the analysis of hydrogen bonds and salt bridges. Similar approaches have been employed in previous studies to optimize the binding interfaces of vaccine-receptor complexes, leading to improved affinity of vaccine candidates to the receptor [81,86]. MD simulations have provided valuable insights into the dynamic behavior of vaccine-TLR complexes over time. The analyses of RMSD and Rg values proved the stability of these complexes. Additionally, RMSF analysis revealed specific regions within the vaccine candidate exhibiting flexibility. Hydrogen bond analysis and Gibbs free energy analysis results corroborated with RMSD and Rg, provided evidence for the stability of the binding between the vaccine and TLRs.These analyses provided theoretical support for the ability of the designed vaccine (WLcd1) to effectively stimulate immune responses.

The analysis of population coverage revealed the potential effectiveness of the WLcd1 vaccine in providing broad protection against CDI, especially in regions with high disease prevalence such as North America and Europe. This indicated that mass vaccination efforts could significantly alleviate the burden of CDI, particularly in countries heavily affected by the disease, such as the United States and Australia. Immune simulation results demonstrated that multiple vaccine doses lead to increased antibody levels, plasma cells, and B cells, indicating a strong humoral response. Additionally, the activation of cytotoxic T-cells and helper T-cells suggested the induction of a potent cellular immune response. These findings revealed the vaccine's potential to stimulate comprehensive immune protection against CDI. Regarding the mRNA vaccine design, the optimized WLcd1-mRNA construct showed favorable characteristics for efficient translation and expression in host cells. In the future, nanoparticles could be used to deliver mRNA to further enhance the stability and immunogenicity of the mRNA vaccine [93].

5. Conclusion

CDI is estimated to affect millions of people worldwide annually, causing significant morbidity, mortality, and healthcare costs. Few studies have been conducted on simultaneously targeting multiple *C. difficile* strains in current *C. difficile* vaccine research. Our study adopts innovative and well-founded approaches, starting with the *C. difficile* pan-genome and target antigen conservation, proceeding to the restriction of HLA alleles, and finally assessing the affinity between vaccine candidates and TLRs. This comprehensive strategy aims to enhance the immunogenicity and applicability of the *C. difficile* vaccine. Given the diversity of *C. difficile* strains and the pivotal role of CTL immunity in antimicrobial defense, the significance of the WLcd1 mRNA vaccine surpasses that of previous research efforts. In summary, the mRNA vaccine has the potential to combat various *C. difficile* strains, but its efficacy and safety need to be verified through further experiments.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Studies in humans and animals

Not applicable.

Availability of data and materials

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Funding

CRediT authorship contribution statement

Caixia Tan: Conceptualization, Methodology, Software, Data curation, Writing – original draft, preparation. **Yuanyuan xiao:** Writing – original draft, preparation. **Ting Liu:** Writing – original draft, preparation. **Siyao Chen:** Writing – original draft, preparation. **Juan Zhou:** Writing – original draft, preparation. **Sisi Zhang:** Writing – original draft, preparation. **Yiran Hu:** Writing – original draft, All authors have read and agreed to the published version of the manuscript. **Anhua Wu:** Conceptualization, Writing- Reviewing and Editing. **Chunhui Li:** Conceptualization, Writing- Reviewing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2024.05.008.

References

[1] Leffler DA, Lamont JT. Clostridium difficile infection. N Engl J Med 2015;372: 1539–48.
[2] Guh AY, Mu Y, Winston LG, Johnston H, Olson D, Farley MM, et al. Trends in U.S. Burden of Clostridioides difficile infection and outcomes. N Engl J Med 2020;382: 1320–30.
[3] European Centre for Disease Prevention and Control. Clostridioides (Clostridium) difficile infections - annual epidemiological report for 2016–2017. https://www.ecdc.europa.eu/en/publications-data/clostridiodes-difficile-infections-annual-e pidemiological-report-2016-2017. [Accessed 20 June 2023].
[4] Doll M, Marra AR, Apisarnthanarak A, Al-Maani AS, Abbas S, Rosenthal VD. Prevention of Clostridioides difficile in hospitals: a position paper of the international society for infectious diseases. Int J Infect Dis 2021;102:188–95.
[5] Guery B, Galperine T, Barbut F. Clostridioides difficile: diagnosis and treatments. Bmj 2019;366:l4609.
[6] Song JH, Kim YS. Recurrent Clostridium difficile infection: risk factors, treatment, and prevention. Gut Liver 2019;13:16–24.
[7] Mahapatra SR, Dey J, Raj TK, Misra N, Suar M. Designing a next-generation multiepitope-based vaccine against Staphylococcus aureus using reverse vaccinology approaches. Pathogens 2023;12.
[8] Chukwudozie OS, Duru VC, Ndiribe CC, Aborode AT, Oyebanji VO, Emikpe BO. The relevance of bioinformatics applications in the discovery of vaccine candidates and potential drugs for COVID-19 treatment. Bioinf Biol Insights 2021;15: 11779322211002168.
[9] Hegde NR, Gauthami S, Sampath Kumar HM, Bayry J. The use of databases, data mining and immunoinformatics in vaccinology: where are we? Expet Opin Drug Discov 2018;13:117–30.
[10] Rcheulishvili N, Mao J, Papukashvili D, Liu C, Wang Z, Zhao J, et al. Designing multi-epitope mRNA construct as a universal influenza vaccine candidate for future epidemic/pandemic preparedness. Int J Biol Macromol 2023;226:885–99.
[11] Hou F, Zhang Y, Liu X, Murad YM, Xu J, Yu Z, et al. mRNA vaccines encoding fusion proteins of monkeypox virus antigens protect mice from vaccinia virus challenge. Nat Commun 2023;14:5925.
[12] Verbeke R, Hogan MJ, Loré K, Pardi N. Innate immune mechanisms of mRNA vaccines. Immunity 2022;55:1993–2005.
[13] To KKW, Cho WCS. An overview of rational design of mRNA-based therapeutics and vaccines. Expet Opin Drug Discov 2021;16:1307–17.
[14] Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines - a new era in vaccinology. Nat Rev Drug Discov 2018;17:261–79.
[15] Chaudhari NM, Gupta VK, Dutta C. BPGA- an ultra-fast pan-genome analysis pipeline. Sci Rep 2016;6:24373.
[16] Luo H, Lin Y, Gao F, Zhang CT, Zhang R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. Nucleic Acids Res 2014;42:D574–80.
[17] Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res 2022;50:D912–d917.
[18] Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 2020;48:D517–d525.
[19] Sudeshna Panda S, Dey J, Mahapatra SR, Kushwaha GS, Misra N, Suar M, et al. Investigation on structural prediction of pectate lyase enzymes from different microbes and comparative docking studies with pectin: the economical waste from food industry. Geomicrobiol J 2021;39:294–305.
[20] Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 2010;26:1608–15.
[21] Chou K-C, Shen H-B. Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms. Nat Protoc 2008;3:153–62.
[22] Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinf 2007;8:4.
[23] Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A, Felgner PL, et al. High-throughput prediction of protein antigenicity using protein microarray data. Bioinformatics 2010;26:2936–43.
[24] Sharma N, Naorem LD, Jain S, Raghava GPS. ToxinPred2: an improved method for predicting toxicity of proteins. Briefings Bioinf 2022;23.
[25] Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2–a server for in silico prediction of allergens. J Mol Model 2014;20:2278.
[26] H. Jeppe, D.T. Konstantinos, P. Mads Damgaard, A. José Juan Almagro, M. Paolo, N. Henrik, et al., DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks, bioRxiv, DOI 10.1101/2022.04.08.487609 (2022).2022.2004.2008.487609.

[27] Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 2022;40:1023–5.

[28] Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. The proteomics protocols handbook. Totowa, NJ: Humana Press; 2005. p. 571–607.

[29] Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinf 2007;8:424.

[30] Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP. In silico approach for predicting toxicity of peptides and proteins. PLoS One 2013;8: e73957.

[31] Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. PLoS Comput Biol 2013;9:e1003266.

[32] Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res 2020;48:W449–w454.

[33] Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 2008;4:e1000048.

[34] Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, et al. Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinf 2010;11:568.

[35] Ivashkiv LB. IFNγ: signalling, epigenetics and roles in immunity, metabolism, disease and cancer immunotherapy. Nat Rev Immunol 2018;18:545–58.

[36] Heeb LEM, Egholm C, Boyman O. Evolution and function of interleukin-4 receptor signaling in adaptive immunity and neutrophils. Gene Immun 2020;21:143–9.

[37] Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. Biol Direct 2013;8:30.

[38] Dhanda SK, Gupta S, Vir P, Raghava GP. Prediction of IL4 inducing peptides. Clin Dev Immunol 2013;2013:263952.

[39] Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 2006;65:40–8.

[40] Zhao T, Cai Y, Jiang Y, He X, Wei Y, Yu Y, et al. Vaccine adjuvants: mechanisms and platforms. Signal Transduct Targeted Ther 2023;8:283.

[41] Shelley JR, Davidson DJ, Dorin JR. The dichotomous responses driven by β-defensins. Front Immunol 2020;11:1176.

[42] Shanmugam A, Rajoria S, George AL, Mittelman A, Suriano R, Tiwari RK. Synthetic Toll like receptor-4 (TLR-4) agonist peptides as a novel class of adjuvants. PLoS One 2012;7:e30839.

[43] Gu Y, Yang J, He C, Zhao T, Lu R, Liu J, et al. Incorporation of a Toll-like receptor 2/6 agonist potentiates mRNA vaccines against cancer and infectious diseases. Signal Transduct Targeted Ther 2023;8:273.

[44] Kaur A, Kaushik D, Piplani S, Mehta SK, Petrovsky N, Salunke DB. TLR2 agonistic small molecules: detailed structure-activity relationship, applications, and future prospects. J Med Chem 2021;64:233–78.

[45] Ghaffari-Nazari H, Tavakkol-Afshari J, Jaafari MR, Tahaghoghi-Hajghorbani S, Masoumi E, Jalali SA. Improving multi-epitope long peptide vaccine potency by using a strategy that enhances CD4+ T help in BALB/c mice. PLoS One 2015;10: e0142563.

[46] Sanches RCO, Tiwari S, Ferreira LCG, Oliveira FM, Lopes MD, Passos MJF, et al. Immunoinformatics design of multi-epitope peptide-based vaccine against schistosoma mansoni using transmembrane proteins as a target. Front Immunol 2021;12:621706.

[47] Kolla HB, Tirumalasetty C, Sreerama K, Ayyagari VS. An immunoinformatics approach for the design of a multi-epitope vaccine targeting super antigen TSST-1 of Staphylococcus aureus. J Genet Eng Biotechnol 2021;19:69.

[48] Livingston B, Crimi C, Newman M, Higashimoto Y, Appella E, Sidney J, and A. A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. J Immunol 2002;168:5499–506.

[49] Yano A, Onozuka A, Asahi-Ozaki Y, Imai S, Hanada N, Miwa Y, et al. An ingenious design for peptide vaccine. Vaccine 2005;23:2322–6.

[50] Buchan DWA, Jones DT. The PSIPRED protein analysis workbench: 20 years on. Nucleic Acids Res 2019;47:W402–w407.

[51] Chatterjee R, Mahapatra SR, Dey J, Raj Takur K, Raina V, Misra N, et al. An immunoinformatics and structural vaccinology study to design a multi-epitope vaccine against Staphylococcus aureus infection. J Mol Recogn 2023;36:e3007.

[52] Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 2004;32:W526–31.

[53] Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. Nucleic Acids Res 2012;40:W294–7.

[54] Singh A, Kaushik R, Jayaram B. Quality assessment of protein tertiary structures: past, present, and future. In: Shanker A, editor. Bioinformatics: sequences, structures, phylogeny. Singapore: Springer Singapore; 2018. p. 271–88.

[55] Mahapatra SR, Dey J, Raj TK, Kumar V, Ghosh M, Verma KK, et al. The potential of plant-derived secondary metabolites as novel drug candidates against Klebsiella pneumoniae: molecular docking and simulation investigation. South Afr J Bot 2022;149:789–97.

[56] Narang PK, Dey J, Mahapatra SR, Roy R, Kushwaha GS, Misra N, et al. Genome-based identification and comparative analysis of enzymes for carotenoid biosynthesis in microalgae. World J Microbiol Biotechnol 2021;38:8.

[57] Narang PK, Dey J, Mahapatra SR, Ghosh M, Misra N, Suar M, et al. Functional annotation and sequence-structure characterization of a hypothetical protein putatively involved in carotenoid biosynthesis in microalgae. South Afr J Bot 2021; 141:219–26.

[58] Vijay K. Toll-like receptors in immunity and inflammatory diseases: past, present, and future. Int Immunopharm 2018;59:391–412.

[59] Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. Nat Protoc 2017;12:255–78.

[60] van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. J Mol Biol 2016;428:720–5.

[61] Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: structural summaries of PDB entries. Protein Sci 2018;27:129–34.

[62] Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci 2021;30:70–82.

[63] Kutzner C, Páll S, Fechner M, Esztermann A, de Groot BL, Grubmüller H. Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. J Comput Chem 2015;36:1990–2008.

[64] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theor Comput 2015;11:3696–713.

[65] Tan S, Do DD, Nicholson D. Consistency of NVT, NPT, μVT and Gibbs (NV2T and NPT) with kinetic Monte Carlo schemes. Chem Eng J 2020;401:126056.

[66] Dey J, Mahapatra SR, Singh PK, Prabhuswamimath SC, Misra N, Suar M. Designing of multi-epitope peptide vaccine against Acinetobacter baumannii through combined immunoinformatics and protein interaction–based approaches. Immunol Res 2023;71:639–62.

[67] Dey J, Mahapatra SR, Raj TK, Kaur T, Jain P, Tiwari A, et al. Designing a novel multi-epitope vaccine to evoke a robust immune response against pathogenic multidrug-resistant Enterococcus faecium bacterium. Gut Pathog 2022;14:21.

[68] Maisuradze GG, Liwo A, Scheraga HA. Relation between free energy landscapes of proteins and dynamics. J Chem Theor Comput 2010;6:583–95.

[69] Dey J, Mahapatra SR, Raj TK, Misra N, Suar M. Identification of potential flavonoid compounds as antibacterial therapeutics against Klebsiella pneumoniae infection using structure-based virtual screening and molecular dynamics simulation. Mol Divers 2023. https://doi.org/10.1007/s11030-023-10738-z.

[70] Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E. gmx_MMPBSA: a new tool to perform end-state free energy calculations with gromacs. J Chem Theor Comput 2021;17:6281–91.

[71] Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinf 2006;7:153.

[72] Rapin N, Lund O, Castiglione F. Immune system simulation online. Bioinformatics 2011;27:2013–4.

[73] Mahapatra SR, Dey J, Jaiswal A, Roy R, Misra N, Suar M. Immunoinformatics-guided designing of epitope-based subunit vaccine from Pilus assembly protein of Acinetobacter baumannii bacteria. J Immunol Methods 2022;508:113325.

[74] Dey J, Mahapatra SR, Lata S, Patro S, Misra N, Suar M. Exploring Klebsiella pneumoniae capsule polysaccharide proteins to design multiepitope subunit vaccine to fight against pneumonia. Expert Rev Vaccines 2022;21:569–87.

[75] Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 2005;33:W526–31.

[76] Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. Nucleic Acids Res 2008;36:W70–4.

[77] Fatima R, Aziz M. The hypervirulent strain of Clostridium difficile: NAP1/B1/027 - a brief overview. Cureus 2019;11:e3977.

[78] CDC. About cdc's urgent threats. 2019. https://www.cdc.gov/drugresistance/biggest-threats.html#cdiff. [Accessed 20 June 2023]. 2021.

[79] U.S. National Library of Medicine. A phase 3, randomized observer-blinded study to evaluate the immunogenicity, safety, and tolerability of 2 doses compared to 3 doses of Clostridium difficile vaccine in adults 50 Years of age and older. https://clinicaltrials.gov/ct2/show/NCT03918629. [Accessed 20 June 2023].

[80] Basak S, Deb D, Narsaria U, Kar T, Castiglione F, Sanyal I, et al. In silico designing of vaccine candidate against Clostridium difficile. Sci Rep 2021;11:14215.

[81] Tan C, Zhu F, Xiao Y, Wu Y, Meng X, Liu S, et al. Immunoinformatics approach toward the introduction of a novel multi-epitope vaccine against Clostridium difficile. Front Immunol 2022;13:887061.

[82] Akkaya M, Kwak K, Pierce SK. B cell memory: building two walls of protection against pathogens. Nat Rev Immunol 2020;20:229–38.

[83] Ma S, Zhu F, Xu Y, Wen H, Rao M, Zhang P, et al. Development of a novel multi-epitope mRNA vaccine candidate to combat HMPV virus. Hum Vaccines Immunother 2023;19:2293300.

[84] Tan C, Zhu F, Wu A, Li C. In silico development of a novel anti-mutation, multi-epitope mRNA vaccine against MPXV variants of emerging lineage and sub-lineages by using immunoinformatics approaches. J Biomol Struct Dyn 2024:1–18. https://doi.org/10.1080/07391102.2024.2325109.

[85] Jiang F, Han Y, Liu Y, Xue Y, Cheng P, Xiao L, et al. A comprehensive approach to developing a multi-epitope vaccine against Mycobacterium tuberculosis: from in silico design to in vitro immunization evaluation. Front Immunol 2023;14: 1280299.

[86] Tan C, Zhu F, Pan P, Wu A, Li C. Development of multi-epitope vaccines against the monkeypox virus based on envelope proteins using immunoinformatics approaches. Front Immunol 2023;14:1112816.

[87] Kaushik V, Jain P, Akhtar N, Joshi A, Gupta LR, Grewal RK, et al. Immunoinformatics-aided design and in vivo validation of a peptide-based

multiepitope vaccine targeting canine circovirus. ACS Pharmacol Transl Sci 2022; 5:679–91.

[88] Cheng P, Jiang F, Wang G, Wang J, Xue Y, Wang L, et al. Bioinformatics analysis and consistency verification of a novel tuberculosis vaccine candidate HP13138PB. Front Immunol 2023;14:1102578.

[89] Gu Y, Yang J, He C, Zhao T, Lu R, Liu J, et al. Incorporation of a Toll-like receptor 2/6 agonist potentiates mRNA vaccines against cancer and infectious diseases. Signal Transduct Targeted Ther 2023;8:273.

[90] Yu M, Zhu Y, Li Y, Chen Z, Sha T, Li Z, et al. Design of a novel multi-epitope vaccine against echinococcus granulosus in immunoinformatics. Front Immunol 2021;12: 668492.

[91] Mariuzza RA, Agnihotri P, Orban J. The structural basis of T-cell receptor (TCR) activation: an enduring enigma. J Biol Chem 2020;295:914–25.

[92] Rossjohn J, Gras S, Miles JJ, Turner SJ, Godfrey DI, McCluskey J. T cell antigen receptor recognition of antigen-presenting molecules. Annu Rev Immunol 2015;33: 169–200.

[93] Sahoo P, Dey J, Mahapatra SR, Ghosh A, Jaiswal A, Padhi S, et al. Nanotechnology and COVID-19 convergence: toward new planetary health interventions against the pandemic. OMICS A J Integr Biol 2022;26:473–88.